JOÃO FILIPE LITA DA SILVA

CONSISTÊNCIA FORTE DE ESTIMADORES: O ESTIMADOR DOS MÍNIMOS QUADRADOS

Dissertação apresentada para obtenção do Grau de Doutor em Matemática, na especialidade de Estatística, pela Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia.

ORIENTADOR: PROFESSOR DOUTOR JOÃO TIAGO MEXIA
CO-ORIENTADOR: PROFESSOR DOUTOR PEDRO CORTE REAL

LISBOA
2007

If we turn, turn, turn, turn, turn, turn, turn, turn
And if we turn, turn, turn, turn, turn
Then we might learn, learn

Francis Healy
(Vocalista da banda escocesa TRAVIS)

Agradecimentos

Os meus primeiros agradecimentos vão naturalmente para o Professor Doutor João Tiago Mexia que, desde o primeiro dia, acreditou ser possível a concretização deste meu sonho, não colocando quaisquer obstáculos à minha formação base em análise funcional e equações com derivadas parciais, mas encarando isso como uma eventual vantagem. Devo dizer que este magnífico e entusiasmante desafio me deu um novo ânimo relativamente à matemática, ao ter a possibilidade de estudar estatística e teoria de probabilidades, áreas que para mim constituiam um mundo a descobrir. E neste ponto o Professor Doutor João Tiago Mexia teve um papel absolutamente fulcral ao partilhar comigo o seu vastíssimo conhecimento e a sua experiência de muitos anos nestas áreas, contribuindo para que eu realizasse um enorme salto evolutivo desde Janeiro de 2004 até aos dias de hoje. Gostaria ainda de destacar dois aspectos na personalidade do Professor Doutor João Tiago Mexia que, quanto a mim, foram determinantes para a elaboração deste projecto: por um lado, uma notável e inesgotável capacidade para motivar e incentivar e por outro, sem nunca deixar de me acompanhar, depositou em mim toda a confiança ao dar-me total liberdade na escolha da metodologia de trabalho. Estes factores foram, sem dúvida nenhuma, os ingredientes necessários para que este manuscrito se tornasse uma realidade e que implicaram uma responsabilização pessoal no sentido de corresponder ao crédito em mim depositado.

Quero expressar o meu agradecimento ao Professor Doutor Pedro Corte Real que, nos primeiros tempos, teve sempre disponibilidade e paciência para as questões que constantemente lhe coloquei, algumas delas típicas de quem dá os primeiros passos numa nova área de estudo. Estou também eternamente grato ao Professor Doutor Manuel Leote Esquível que me acompanhou e amparou ao longo deste projecto, muito especialmente no início; o seu conhecimento profundo de excelentes referências bibliográficas veio a revelar-se vital para a concretização deste trabalho.

Uma palavra de enorme apreço ainda para o Professor Doutor Hilmar Drygas, a Professora Doutora Claudia Klüppelberg e o Professor Doutor Alastair Hall que, sem hesitação, me enviaram por correio de superfície alguns dos seus artigos, fundamentais na elaboração desta dissertação.

A concretização de uma tese de doutoramento é sempre o culminar de muitos anos de aprendizagem, razão pela qual seria injusto não destacar todos aqueles que, ao longo desta caminhada, contribuíram directa ou indirectamente para os meus conhecimentos em matemática.

Por último, agradeço todo o apoio e incentivo dados pela minha família sem os quais teria sido impossível a concretização deste sonho, em especial pelos meus Pais que disponibilizaram toda a logística indispensável para um projecto desta envergadura.

Sumário

Neste trabalho, a consistência forte do estimador dos mínimos quadrados em modelos lineares será analisada impondo condições sobre a matriz de modelo e assumindo os mais diversos cenários para os erros aleatórios. De início, serão estabelecidas condições gerais de suficiência para o caso em que os erros são independentes e identicamente distribuídos com variância infinita. Em seguida e utilizando um resultado de convergência quase certa para somas de variáveis aleatórias independentes estáveis será explorada a situação particular de erros independentes com distribuição estável. Esta questão da estabilidade dos erros será ainda estudada com grande detalhe no capítulo seguinte recorrendo, desta feita, a uma importante estimativa sobre a norma da diferença entre o estimador dos mínimos quadrados e o vector parâmetro; daremos especial ênfase ao caso em que os erros são independentes e identicamente distribuídos com distribuição normal onde, à custa desta técnica inovadora, se conseguem estabelecer resultados análogos aos obtidos por Lai, Robbins & Wei nos finais da década de setenta. A estimativa referida anteriormente, servirá também para desenvolver resultados de consistência forte para o estimador dos mínimos quadrados onde não se impõem quaisquer condições de independência ou idêntica distribuição sobre os erros, mas se admite como hipótese a simetria radial do vector dos erros aleatórios. Uma última abordagem a esta temática será realizada, assumindo que a distribuição do quadrado erros aleatórios se encontra numa das três grandes classes: domínio de atracção maximal da distribuição de Fréchet, domínio de atracção maximal da distribuição de Weibull e domínio de atracção maximal da distribuição de Gumbel. A grande novidade aqui reside no facto de usarmos expressamente, na maioria dos resultados, os comportamentos assimptóticos de cauda característicos das distribuições que pertencem a cada um dos três domínios citados previamente.

A finalizar, no último capítulo, será analisada a consistência forte de estimadores não lineares num modelo específico para as observações, onde é introduzida uma metodologia de linearização e se admite que os erros aleatórios satisfazem hipóteses adequadas, nomeadamente, a de possuírem decaimento exponencial no infinito.

Summary

In this work, the strong consistency of least squares estimators in multiple regression models will be studied, given conditions on the design matrix and under several assumptions for the random errors. Starting with establishing general conditions of sufficiency for the strong consistency of the least squares estimators, where it is admitted independence and identical distribution for random errors with infinite variance. Using a result on almost sure convergence for sums of independent stable random variables for stable errors, the situation of stable independent errors will be explored. This situation of stable errors will be studied in detail on the next chapter. There, using an important estimate for the norm of the difference between the least squares estimator and parameter vector, a special emphasis will be given to the case where the random errors are independent and identically distributed with gaussian distribution. Through this new technique, results similar to those obtained by Lai, Robbins & Wei in the seventies will be established. The estimate quoted previously will be very useful in deriving strong consistency results for least squares estimators when assumptions of independence and identical distribution for the random errors are completely excluded and a radial symmetry condition on random vector will be admitted. A last approach to this subject will be carried out assuming that the distribution of random errors is in one of the three classes: maximal domain of attraction of the Fréchet distribution, of the Weibull distribution and of the Gumbel distribution. This new approach uses the specific behaviour of the asymptotic tail of distributions for each of the three domains quoted.

In the last chapter, the strong consistency of nonlinear estimators on a particular model for observations will be analyzed. A linearization methodology will be used and an adequate hypothesis for the random errors will be assumed, namely, that they have exponencial decay at infinity.

Lista de símbolos e abreviaturas

$\mathbb{N}$	conjunto dos números naturais
$\mathbb{R}$, $\mathbb{R}^n$	conjunto dos números reais, espaço euclideano n-dimensional
$\mathbb{S}^{n-1}$	esfera unitária de $\mathbb{R}^n$
$\mathscr{V}_\varepsilon(x_0)$	vizinhança de x_0 com raio ε
$\mathscr{B}_\varepsilon(\boldsymbol{x}_0)$	bola aberta de centro $\boldsymbol{x}_0$ e raio $\varepsilon > 0$
$\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$	σ-álgebras
$\mathcal{B}(\mathbb{R}^n)$	σ-álgebra dos borelianos de $\mathbb{R}^n$
$\pi(\mathbb{R}^n)$	π-sistema de todos os rectângulos de $\mathbb{R}^n$
$\mathrm{Leb}(\,\cdot\,)$	medida de Lebesgue sobre $\mathbb{R}$
$\mathbb{P}(\,\cdot\,)$	medida de probabilidade
q.c.	quase certamente
$(\Omega, \mathcal{F})$	espaço amostral
$(\Omega, \mathcal{F}, \mathbb{P})$	espaço de probabilidade
v.a. (v.a.'s)	variável aleatória (variáveis aleatórias)
$\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \ldots$	variáveis aleatórias
$\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$	erros aleatórios
M_n	$\max(\mathrm{X}_1, \ldots, \mathrm{X}_n)$
$\mathrm{X}^+, \mathrm{X}^-$	parte positiva de X, parte negativa de X
$\mathrm{X} \overset{\mathrm{d}}{=} \mathrm{Y}$, $\mathrm{X} \sim \mathrm{Y}$	X e Y identicamente distribuídas
$\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$	Z tem distribuição estável
$\mathrm{X} \sim N(\mu, \sigma^2)$	X tem distribuição normal com média μ e variância σ^2
$\mathrm{X} \sim \chi^2(\nu)$	X tem distribuição χ^2 com ν graus de liberdade
$\mathrm{X} \sim \mathrm{DEI}(\alpha, \beta, \tau)$	X tem decaimento exponencial no infinito
$(\mathrm{X}_1, \ldots, \mathrm{X}_n)$	vector aleatório
i.i.d.	independentes e identicamente distribuídas
$\mathbb{E}(\mathrm{X})$	valor médio de X
$\mathbb{E}\big[\,\lvert\mathrm{X}\rvert^p\,\big]$	momento absoluto de ordem p de X
$\mathbb{V}(\mathrm{X})$	variância de X
$\mathscr{L}_p(\Omega, \mathcal{F}, \mathbb{P})$	espaço $\mathscr{L}_p$
sup q.c.	supremo quase certo
$\mathbb{E}\big\{\mathrm{X} \,\vert\, \mathcal{H}\big\}$	esperança condicional de X dada a σ-álgebra $\mathcal{H}$

I_A	função indicador do conjunto A
φ_{X}	função característica da variável aleatória X
f.d.	função de distribuição
$\mathrm{Supp}(F)$	suporte da função de distribuição F
F_{X}, $\overline{F}_{\mathrm{X}}$	função de distribuição de X, cauda de F_{X}
f_{X}	função densidade de probabilidade de X
$F_{\mathrm{X}_1,\ldots,\mathrm{X}_n}$	função de distribuição conjunta de $(\mathrm{X}_1,\ldots,\mathrm{X}_n)$
$F^{\leftarrow}$	função quantil
$\mathit{A}(x)$	função auxiliar
$\Phi_\alpha(x)$, $\Lambda(x)$, $\Psi_\alpha(x)$	distribuições de valores extremos
$\mathcal{D}(\Phi_\alpha)$, $\mathcal{D}(\Lambda)$, $\mathcal{D}(\Psi_\alpha)$	domínio de atracção maximal de Φ_α, Λ_α, Ψ_α
$\longrightarrow$	convergência
$\xrightarrow{\text{q.c.}}$	convergência quase certa
$\xrightarrow{\mathbb{P}}$	convergência em probabilidade
$\xrightarrow{\mathscr{L}_p}$	convergência em momento de ordem p
$\xrightarrow{\mathrm{d}}$	convergência em distribuição ou em lei
$\mathcal{M}_{n\times\kappa}(\mathbb{R})$	conjunto das matrizes rectangulares de $n\times\kappa$ sobre $\mathbb{R}$
$\boldsymbol{I}$, $\mathbf{O}$	matriz identidade, matriz nula
$\boldsymbol{A}^T$, $\boldsymbol{A}^{-1}$	matriz transposta de $\boldsymbol{A}$, matriz inversa de $\boldsymbol{A}$
$\langle\,\cdot\,,\,\cdot\,\rangle$	produto interno canónico
$\boldsymbol{x}\perp\boldsymbol{y}$	$\boldsymbol{x}$ ortogonal a $\boldsymbol{y}$
$\boldsymbol{P}_{\mathbb{F}}\boldsymbol{x}$	projecção ortogonal do vector $\boldsymbol{x}$ sobre $\mathbb{F}$
$\mathrm{car}(\boldsymbol{A})$	característica da matriz $\boldsymbol{A}$
$\mathrm{Nuc}(\boldsymbol{A})$, $\mathrm{Im}(\boldsymbol{A})$	núcleo de $\boldsymbol{A}$, espaço imagem de $\boldsymbol{A}$
$\dim(\mathbb{F})$	dimensão do subespaço vectorial $\mathbb{F}$
$\mathrm{Spec}(\boldsymbol{A})$	espectro da matriz $\boldsymbol{A}$
$\rho(\boldsymbol{A})$	raio espectral da matriz $\boldsymbol{A}$
$\lVert\boldsymbol{x}\rVert$	norma euclideana do vector $\boldsymbol{x}$
$\lVert\boldsymbol{x}\rVert_\infty$	norma do máximo do vector $\boldsymbol{x}$
$\lVert\boldsymbol{A}\rVert$	$\sup_{\boldsymbol{x}\neq\mathbf{0}}\frac{\lVert\boldsymbol{A}\boldsymbol{x}\rVert}{\lVert\boldsymbol{x}\rVert}$
$\lvert\boldsymbol{x}\rvert_{\min}$	$\min\big(\lvert x_1\rvert,\ldots,\lvert x_\kappa\rvert\big)$
$\mathfrak{b}_n$	números de Bernoulli
$\limsup$, $\liminf$, $\lim$	limite superior, limite inferior, limite
$f=O(g)$, $x\to x_0$	f fracamente inferior a g em x_0
$f\asymp g$, $x\to x_0$	f do mesmo grau de g em x_0
$f=o(g)$, $x\to x_0$	f fortemente inferior a g em x_0
$f\sim g$, $x\to x_0$	f (assimptoticamente) equivalente a g em x_0
$\Gamma(x)$	função gama
$\Gamma(x,y)$	função gama incompleta

$\mathrm{erf}(x)$	função erro
$\mathcal{B}_n(x)$	função de Bessel (de 1ª espécie)
$\mathrm{K}_n(x)$	função de Bessel modificada
$\mathfrak{l}_k$	integral de Euler-Mascheroni
$\log_m x$	$\log_1 x = \log(x)$, $\log_2 x = \log(\log(x)), \ldots$
$\exp_m x$	$\exp_1 x = \exp(x)$, $\exp_2 x = \exp(\exp(x)), \ldots$
$[x]$	maior inteiro menor ou igual a x
δ_{ip}	símbolo de Kronecker
$(\alpha)_n$	símbolo de Pochhammer
${}_p\mathrm{F}_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x)$	função hipergeométrica generalizada
${}_p\widetilde{\mathrm{F}}_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x)$	função hipergeométrica generalizada regularizada
$\overline{D}$	aderência de D
$\max\limits_{1\leqslant j\leqslant n} a_j$	$\max(a_1, \ldots, a_n)$
$\max\limits_{\mathbb{H}} f$	máximo da função f no conjunto $\mathbb{H}$
$\mathrm{grad}(f)$	gradiente de f
$\mathrm{sign}(t)$	função sinal
x_{inf}, x_{sup}	extremo esquerdo do suporte, extremo direito do suporte
$\mathcal{R}_0, \mathcal{R}_\alpha, \mathcal{R}_{-\infty}$	conjunto das funções de variação lenta, regular, rápida no infinito
$\ell^{\#}$	conjugada de Bruijn da função de variação lenta ℓ
EMQ	estimador dos mínimos quadrados
$\mathbf{Y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \mathbf{e}_n$	modelo de regressão linear
$\boldsymbol{X}_n$	matriz de modelo
$\boldsymbol{\beta}$	vector parâmetro
$\widetilde{\boldsymbol{\beta}}$	estimador dos mínimos quadrados
$\mathbf{e}_n$	vector dos erros
$\blacksquare$	fim de demonstração

Índice

Prefácio

O principal objectivo do presente trabalho é estabelecer a consistência forte do estimador dos mínimos quadrados em modelos lineares assumindo as mais diversas hipóteses para os erros aleatórios e impondo condições sobre a matriz de modelo.

No primeiro capítulo, serão recordadas algumas noções elementares de teoria de probabilidades, distribuições estáveis unidimensionais e teoria de valores extremos expondo-se, por um lado, um conjunto de definições que servirão de base para os vários capítulos posteriores e, por outro, uma colectânea de úteis resultados com importância decisiva no conteúdo original. De destacar que na segunda e terceira secções do primeiro capítulo serão relembrados também alguns conceitos de análise matricial e real, respectivamente, que funcionaram como ferramentas auxiliares ao longo de todo o texto. O primeiro capítulo é finalizado com uma última secção dedicada exclusivamente à análise de regressão dando especial destaque à motivação que conduz ao modelo de regressão linear.

No segundo capítulo, serão anunciados os primeiros resultados originais que constituem uma extensão das condições de suficiência para a consistência forte do estimador dos mínimos quadrados, estabelecidas na década de noventa por Mingzhong Jin e Xiru Chen, ao caso em que os erros aleatórios do modelo linear não possuem momentos absolutos. Ainda neste capítulo e seguindo um trabalho realizado para o XIV Congresso Anual da SPE será elaborada uma abordagem alternativa à anterior, assumindo a independência e estabilidade dos erros aleatórios do modelo linear.

No terceiro capítulo e ainda dentro da temática da estabilidade, será estuda com grande detalhe a consistência forte e a consistência em momento de ordem $0 < r \leqslant 2$ do estimador dos mínimos quadrados por um processo de estimativa. A estimativa em causa será obtida sobre a norma da diferença entre o estimador dos mínimos quadrados e o vector parâmetro recorrendo apenas a operações de cálculo matricial. Através desta poderosa técnica, conseguem-se estabelecer resultados análogos aos obtidos por Lai, Robbins & Wei nos finais da década de setenta num cenário em que os erros são independentes e identicamente distribuídos com distribuição normal. Usufruindo ainda das magníficas propriedades das distribuições estáveis, conseguem-se generalizar os resultados do caso normal à situação em que a variância dos erros não está definida assumindo apenas a independência e a estabilidade destas variáveis aleatórias.

No quarto capítulo, será analisada a consistência forte e a consistência em momento de ordem $s > 0$ do estimador dos mínimos quadrados excluindo quaisquer hipóteses de independência ou idêntica distribuição para os erros aleatórios mas impondo, em alternativa, uma condição de simetria radial sobre o vector dos erros. Recorrendo a um método de factorização de variáveis aleatórias e a coordenadas polares generalizadas consegue-se uma elegante abordagem geométrica à consistência do estimador dos mínimos quadrados num leque de hipóteses pouco habitual. No final, alguns exemplos práticos serão expostos com a respectiva aplicação da teoria desenvolvida anteriormente.

No quinto capítulo e tendo por base um trabalho realizado no XIII Congresso Anual da SPE, será feito um último estudo à consistência forte e em média quadrática do estimador dos mínimos quadrados admitindo que a distribuição do quadrado dos erros aleatórios se encontra numa das três grandes classes: domínio de atracção maximal da distribuição de Fréchet, domínio de atracção maximal da distribuição de Weibull e domínio de atracção maximal da distribuição de Gumbel. A grande novidade aqui residirá no facto de utilizarmos, na maioria dos resultados, o comportamento assimptótico específico das distribuições que pertencem a cada um dos três domínios previamente citados.

Finalmente, no último capítulo será analisada a consistência forte de estimadores não-lineares num modelo concreto para as observações, onde se desenvolverá uma metodologia de linearização admitindo que os erros aleatórios satisfazem hipóteses adequadas, em particular, que possuem decaimento exponencial no infinito. O caso em que os erros são independentes e identicamente distribuídos com distribuição normal é estudado numa secção à parte; a terminar serão exibidos alguns exemplos práticos sendo feita a aplicação dos respectivos resultados teóricos.

Faro, Fevereiro de 2007,
João Filipe Lita da Silva

Capítulo 1

Preliminares

1.1 Teoria de probabilidades

Historicamente, a teoria de probabilidades teve a sua origem em jogos de azar e apostas, que gozavam de enorme popularidade no século XVII. As primeiras tentativas para formalizar as ideias intuitivas associadas a este tipo de disputas, devem-se aos famosos matemáticos franceses Blaise Pascal e Pierre de Fermat. Ao que parece, no longínquo ano de 1654, o nobre francês Antoine Gombaud (Chevalier de Méré) chamou a atenção de Pascal para uma aparente contradição num "jogo de dados". O jogo consistia em lançar um par de dados 24 vezes e o problema estava em decidir se era correcto apostar a mesma importância a favor ou contra o aparecimento de pelo menos um "duplo seis", nos vinte e quatro lançamentos. Este e outros problemas colocados por Gombaud motivaram uma troca de correspondência entre Pascal e Fermat, na qual se estabeleceram os princípios fundamentais desta teoria. Um professor de Leibniz, Christian Huygens, conhecedor do conteúdo dessa correspondência publicou em 1657 o livro *De Ratiociniis in Ludo Aleae*, um tratado com problemas associados aos jogos de azar e desde então, o desenvolvimento prosseguiu com o contributo de Jakob Bernoulli e Abraham De Moivre. Em 1812, Pierre Simon Laplace introduziu, na sua *Théorie Analytique des Probabilités*, um grande número de conceitos e técnicas matemáticas, demonstrando a sua aplicação em muitos problemas científicos e práticos (a teoria dos erros, a matemática actuarial e a mecânica estatística foram algumas das áreas que se desenvolveram no século XIX à custa das ideias apresentadas por Laplace). Em particular, formulou o que se costuma designar por **definição clássica de probabilidade**: a probabilidade de um acontecimento é o quociente do número de casos *favoráveis* a esse acontecimento sobre o número de casos *possíveis*, desde que estes sejam equiprováveis (ou igualmente prováveis). Contudo, esta definição para além de restritiva é inadequada já que não define realmente o conceito de probabilidade, sendo apenas um método prático de cálculo da probabilidade de acontecimentos simples, limitado aos conjuntos finitos de todos os resultados possíveis. Efectivamente, o grande problema residia no facto de se poder definir "ao acaso"da maneira que quiséssemos e as diferentes definições conduziam-nos a diferentes respostas. Em 1889, Joseph Bertrand editou *Calcul de Probabilités* expondo um variado número de problemas geométricos onde o resultado dependia do método de solução. Tornou-se, então, imperioso criar uma definição de probabilidade geral que ultrapassasse todas estas dificuldades.

O assunto foi definitivamente resolvido já no século XX, no ano de 1933, por um nome incontornável na teoria de probabilidades: Andrei Nikolaevitch Kolmogorov. Este matemático russo axiomatizou o conceito de probabilidade na sua monografia *Foundations of the Theory of Probability* lançando, assim, bases sólidas para a moderna teoria de probabilidades. Desde então, as ideias de A. N. Kolmogorov foram sendo refinadas, sempre sob o estimulo das aplicações (a estatística é um dos exemplos notáveis para além de outros tão diversificados como a economia, psicologia, genética ou engenharia) e a teoria de probabilidades é, hoje em dia, parte de uma teoria mais geral chamada teoria da medida.

A palavra *probabilidade* de utilização comum no vocabulário quotidiano, está associada a situações em que o conhecimento, no momento, não nos permite saber a evolução futura, isto é, sabemos que há vários resultados possíveis mas desconhecemos qual deles se irá realizar. A este tipo de fenómenos onde o estudo torna indispensável o uso de probabilidades chamamos, em certas condições, experiência aleatória.

Definição 1.1.1 Uma **experiência aleatória** é uma experiência onde:

(E1) Conhecemos todos os seus possíveis resultados.

(E2) Não se conhece antecipadamente qual dos resultados possíveis vai ocorrer cada vez que é efectuada.

(E3) A sua repetição se pode realizar em condições análogas.

Em teoria de probabilidades é estudada a incerteza de uma experiência aleatória, pelo que é conveniente associar a cada experiência, um conjunto Ω de resultados possíveis (i.e. de resultados elementares e indivisíveis) dessa experiência e um outro conjunto que represente todos os *acontecimentos* possíveis da experiência. Mais concretamente, a próxima definição vai ao encontro daquilo que verdadeiramente necessitamos.

Definição 1.1.2 Uma classe $\mathcal{F}$ de subconjuntos de um conjunto não-vazio Ω que satisfaça:

(σ1) $\Omega \in \mathcal{F}$.

(σ2) Se $A \in \mathcal{F}$ então $\Omega \setminus A \in \mathcal{F}$.

(σ3) Se $A_n \in \mathcal{F}$ para $n = 1, 2, \ldots$ então $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

é chamada **σ-álgebra** sobre Ω (ou σ-álgebra de subconjuntos de Ω).

Dada uma classe $\mathcal{C}$ de subconjuntos de um conjunto Ω chamamos **σ-álgebra gerada por $\mathcal{C}$** à menor σ-álgebra sobre Ω que contém $\mathcal{C}$ e indicamos esse facto escrevendo $\sigma(\mathcal{C})$. Note-se, então, que a σ-álgebra gerada por $\mathcal{C}$ será a intersecção de todas as σ-álgebras sobre Ω que contêm $\mathcal{C}$.

Exemplo 1.1.1
Se X for um espaço topológico, chamamos **σ-álgebra dos borelianos** de X à σ-álgebra sobre X gerada pela classe dos subconjuntos abertos de X. Esta σ-álgebra é representada usualmente por $\mathcal{B}(X)$ e os seus elementos têm o nome de **borelianos** de X. Um caso particular importantíssimo acontece quando $X = \mathbb{R}^n$.

A σ-álgebra dos borelianos de $\mathbb{R}^n$ é, sem dúvida, a mais importante de todas as σ-álgebras mas os seus elementos podem ser subconjuntos de $\mathbb{R}^n$ bastante complicados. Neste sentido, a definição e o exemplo que se seguem têm por objectivo clarificar um pouco mais este assunto.

Definição 1.1.3 Sendo Ω um conjunto chamamos **π-sistema** sobre Ω a toda a família $\mathcal{I}$ de subconjuntos de Ω fechada para a intersecção finita,

$$A, B \in \mathcal{I} \quad \Longrightarrow \quad A \cap B \in \mathcal{I}.$$

Exemplo 1.1.2
Uma exemplo típico de um π-sistema é a classe formada por todos os rectângulos de $\mathbb{R}^n$ com a forma,

$$\big\{(x_1, \ldots, x_n) \colon\; x_1 \leqslant a_1, \ldots, x_n \leqslant a_n,\; a_i \in \mathbb{R},\; i = 1, \ldots, n\big\}$$

e designada habitualmente por $\pi(\mathbb{R}^n)$. Além disso, a σ-álgebra dos borelianos de $\mathbb{R}^n$ coincide com a σ-álgebra gerada por $\pi(\mathbb{R}^n)$, ou seja, $\mathcal{B}(\mathbb{R}^n) = \sigma(\pi(\mathbb{R}^n))$.

A ideia de associar, a uma experiência aleatória, um conjunto de todos os resultados possíveis e a respectiva classe de subconjuntos que identifique todos os *acontecimentos* possíveis da experiência pode agora ser formalizada.

Definição 1.1.4 O **espaço amostral** de uma experiência aleatória é um par $(\Omega, \mathcal{F})$ onde Ω é o conjunto de todos os possíveis resultados da experiência e $\mathcal{F}$ uma σ-álgebra de subconjuntos de Ω.

Os elementos de $\mathcal{F}$ designam-se por **acontecimentos**; em particular, quando $\omega \in \Omega$ chama-se **acontecimento elementar** a $\{\omega\}$. Na linguagem da teoria da medida, o par $(\Omega, \mathcal{F})$ da Definição 1.1.4 designa-se por **espaço mensurável** e os elementos de $\mathcal{F}$ são chamados de **$\mathcal{F}$-mensuráveis** (ou simplesmente mensuráveis).

Sendo Ω um conjunto, chamamos **função de conjunto** a qualquer função com valores reais definida numa classe de subconjuntos de Ω. A próxima definição, apresenta os três axiomas fundamentais para a noção de probabilidade.

Definição 1.1.5 Seja $(\Omega, \mathcal{F})$ um espaço amostral. Uma **medida de probabilidade** é uma função de conjunto $\mathbb{P}(\,\cdot\,)$ definida em $\mathcal{F}$ que satisfaz os axiomas:

(P1) $\mathbb{P}(A) \geqslant 0$ para todo o $A \in \mathcal{F}$.

(P2) Se $\{A_n,\, n \geqslant 1\}$ é uma sucessão de elementos de $\mathcal{F}$ mutuamente exclusivos (isto é, $A_i \cap A_j = \emptyset$ se $i \neq j$) então $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

(P3) $\mathbb{P}(\Omega) = 1$.

O termo "medida"na definição acima advém do facto de, no quadro da teoria da medida, dado um espaço mensurável $(\Omega, \mathcal{F})$ chama-se **medida** sobre $(\Omega, \mathcal{F})$ a uma aplicação $\mathbb{m}\colon \mathcal{F} \to [0, \infty]$ que verifica $\mathbb{m}(\emptyset) = 0$ e o axioma (P2). O triplo $(\Omega, \mathcal{F}, \mathbb{m})$ diz-se então um **espaço de medida**.

Exemplo 1.1.3
Sendo X um conjunto qualquer, $(X, \mathcal{F})$ um espaço mensurável e a um elemento fixado de X podemos definir em $\mathcal{F}$ uma medida $\mathbb{m}$ por: $\mathbb{m}(A) = 0$ se $a \notin A$ e $\mathbb{m}(A) = 1$ se $a \in A$, para todo o $A \in \mathcal{F}$. Esta é a **medida de Dirac** relativa ao ponto a.

Exemplo 1.1.4
Um espaço de medida de grande relevo é $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{Leb})$ onde Leb designa a **medida de Lebesgue** sobre $\mathbb{R}$. Considerando a classe $\mathcal{S}$ de partes de $\mathbb{R}$ constituída por todos os intervalos semi-abertos da forma $[a, b[$ $(a \leqslant b)$[1] então $\mathrm{Leb}\colon \mathcal{B}(\mathbb{R}) \longrightarrow [0, \infty]$ é a (única) medida sobre a σ-álgebra dos borelianos de $\mathbb{R}$ que prolonga $\mathrm{Leb}\colon \mathcal{S} \longrightarrow [0, \infty]$ dada por $\mathrm{Leb}([a, b[) = b - a$.

Neste momento, estamos em condições de anunciar a formulação do modelo matemático para uma experiência aleatória, ou seja, o **modelo probabilístico** que servirá de base para todos os estudos que efectuarmos e que é constituído por:

M1. Um conjunto não-vazio Ω, de todos os resultados possíveis.

M2. Uma σ-álgebra $\mathcal{F}$ sobre Ω de acontecimentos aleatórios.

M3. Uma medida de probabilidade $\mathbb{P}$ definida em $\mathcal{F}$.

Enfim, num conceito matemático abstracto tem-se a,

Definição 1.1.6 Um **espaço de probabilidade** é um triplo $(\Omega, \mathcal{F}, \mathbb{P})$ onde Ω é um conjunto não-vazio, $\mathcal{F}$ é uma σ-álgebra de subconjuntos de Ω e $\mathbb{P}$ é uma medida de probabilidade definida em $\mathcal{F}$.

Num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ dizemos que um acontecimento $A \in \mathcal{F}$ ocorre **quase certamente** (abreviadamente q.c.) se $\mathbb{P}(A) = 1$.

---

[1] Uma classe $\mathcal{C}$ de partes de um conjunto Ω diz-se um **semi-anel** se se verificarem as propriedades seguintes:

(S1) $\emptyset \in \mathcal{C}$.

(S2) Se $A \in \mathcal{C}$ e $B \in \mathcal{C}$ então $A \cap B \in \mathcal{C}$.

(S3) Se $A \in \mathcal{C}$ e $B \in \mathcal{C}$ então existe um conjunto finito de índices I e uma família $\{A_i,\, i \in I\}$ de subconjuntos pertencentes a $\mathcal{S}$ disjuntos dois a dois, tal que $A - B = \bigcup_{i \in I} A_i$.

A classe $\mathcal{S}$ de partes de $\mathbb{R}$ formada por todos os intervalos semi-abertos da forma $[a, b[$ $(a \leqslant b)$ constitui um semi-anel. De referir que, o conceito de medida também pode ser definido em qualquer semi-anel (ver [58]).

Seja $\{A_n,\, n \geqslant 1\}$ uma sucessão de acontecimentos num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Definimos **limite superior** da sucessão $\{A_n,\, n \geqslant 1\}$ por,

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

O **limite inferior** da sucessão $\{A_n,\, n \geqslant 1\}$ é definido por,

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m.$$

O acontecimento $\limsup A_n$ é o acontecimento "ocorrência de um número infinito dos A_n", no sentido em que $\omega \in \limsup A_n$ se e só se ω pertence a um número infinito dos A_n. Uma notação alternativa para $\limsup A_n$ é $\{A_n,\ \text{infinitas vezes}\}$. O acontecimento $\liminf A_n$ também tem uma interpretação intuitiva: é o acontecimento "ocorrência de A_n para todo o n suficientemente grande". A razão prende-se com o facto de $\omega \in \liminf A_n$ se e só se $\omega \in A_m$ para todo o m suficientemente grande $(m \geqslant n_0(\omega))$. Em contraponto com a notação $\{A_n,\ \text{infinitas vezes}\}$ alguns autores indicam $\liminf A_n$ por $\{A_n,\ \text{quase certamente}\}$ (ver [91]). No seguimento do que acabámos de expor vem a primeira parte do teorema de Borel-Cantelli, um resultado *sine qua non* em teoria de probabilidades e que constitui um instrumento poderoso nas leis de grandes números. Ao longo desta secção, remeteremos as demonstrações dos resultados anunciados para os livros generalistas sobre teoria de probabilidades citados na bibliografia.

Teorema 1.1.1 (primeiro lema de Borel-Cantelli) *Seja $\{A_n,\, n \geqslant 1\}$ uma sucessão de acontecimentos num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ tal que $\sum_{n=1}^{\infty} \mathbb{P}\left(A_n\right) < \infty$ então,*

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0.$$

Demonstração. Consultar [91], página 27. ∎

Vejamos agora como definir variável aleatória. Na prática, os elementos de um espaço amostral podem coisas muito concretas tais como as faces de um dado, as moléculas de um gás ou até seres humanos. Contudo, não são os próprios elementos do espaço amostral que nos interessam mas sim valores numéricos que a eles estão associados. Por exemplo, para um ser humano podemos medir certas características físicas tais como a idade, o peso e a altura ou ainda outras características tais como o rendimento anual, o número de anos de escolaridade. Informalmente, uma variável aleatória será então uma *característica numérica* do resultado de uma experiência aleatória.

Definição 1.1.7 Dado um espaço amostral $(\Omega, \mathcal{F})$ chamamos **variável aleatória** (abreviadamente v.a.) a uma função $\mathrm{X}\colon \Omega \to \mathbb{R}$ cuja imagem inversa de qualquer boreliano de $\mathbb{R}$ é um acontecimento,

$$\mathrm{X}^{-1}(B) = \big\{\omega\colon\ \mathrm{X}(\omega) \in B\big\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

No âmbito da teoria da medida, se $(\Omega, \mathcal{F})$ e $(\Sigma, \mathcal{G})$ forem espaços mensuráveis diz-se que uma função $\mathrm{f}\colon \Omega \to \Sigma$ é **$\mathcal{F}$-mensurável** (ou apenas mensurável) se $\mathrm{f}^{-1}(A) = \{\omega \in \Omega\colon\ \mathrm{f}(\omega) \in A\} \in \mathcal{F}$ para cada $A \in \mathcal{G}$ (fazendo a analogia com a Definição 1.1.7 conclui-se que uma v.a. não é mais do que uma função $\mathcal{F}$-mensurável cujo o espaço mensurável $(\Sigma, \mathcal{G})$ é $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Exemplo 1.1.5
Se X é uma v.a. definida em $(\Omega, \mathcal{F})$ e $\mathrm{fi}\colon \mathbb{R} \to \mathbb{R}$ uma **função boreliana** (i.e. $\mathcal{B}(\mathbb{R})$-mensurável quando $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R})) = (\Sigma, \mathcal{G})$) então fi(X) é ainda uma v.a. (em particular, se fi for contínua ou monótona então é boreliana).

Se $\{\mathrm{Y}_j,\ j \in J\}$ for uma família não-vazia de v.a.'s sobre $(\Omega, \mathcal{F})$ tem lugar uma σ-álgebra de enorme relevância: a σ-álgebra gerada pela classe de conjuntos $\{\omega\colon\ \mathrm{Y}_j(\omega) \in B\}$, $B \in \mathcal{B}(\mathbb{R})$, $j \in J$ designada habitualmente por $\sigma(\mathrm{Y}_j,\ j \in J)$. Claramente, $\sigma(\mathrm{Y}_j,\ j \in J) \subset \mathcal{F}$ sendo também a menor σ-álgebra relativamente à qual todos os Y_j, $j \in J$ são mensuráveis.

No conceito de v.a. definido anteriormente, não encontramos qualquer referência a uma eventual medida de probabilidade $\mathbb{P}$ definida no espaço amostral. Porém, na prática, as variáveis aleatórias que interessam verdadeiramente encontram-se definidas num espaço de probabilidade e estão, como iremos ver, intimamente ligadas a uma classe especial de funções reais de variável real, a ser introduzida já de seguida.

Definição 1.1.8 Uma função real de variável real F definida em todo o $\mathbb{R}$ que seja não-decrescente[2], contínua à direita[3] e satisfaça,

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{e} \quad \lim_{x \to +\infty} F(x) = 1$$

é designada por **função de distribuição** (abreviadamente f.d.).

O **suporte** de uma f.d. arbitrária F é o conjunto fechado,

$$\mathrm{Supp}(F) = \{x\colon\ F(x+\varepsilon) - F(x-\varepsilon) > 0,\ \forall \varepsilon > 0\}$$

(ver [18] ou [19]). Os elementos de $\mathrm{Supp}(F)$ são chamados **pontos de crescimento** de F. Uma f.d. F diz-se **degenerada** se possui apenas um único ponto de crescimento; caso contrário dizemos que F é **não-degenerada**. Vamos agora introduzir algumas notações que usaremos ao longo de todo o texto. Representaremos a **cauda** de uma f.d. F por,

$$\overline{F}(x) = 1 - F(x), \quad x \geqslant 0.$$

[2]Recordemos que uma função F definida em $D \subset \mathbb{R}$ é não-decrescente (resp. não-crescente) se dados $x, y \in D$, $x < y \implies F(x) \leqslant F(y)$ (resp. $x, y \in D$, $x < y \implies F(x) \geqslant F(y)$). Enfim, uma função F definida em $D \subset \mathbb{R}$ é crescente (resp. decrescente) se dados $x, y \in D$, $x < y \implies F(x) < F(y)$ (resp. $x, y \in D$, $x < y \implies F(x) > F(y)$).

[3]Lembremos que uma função F definida em todo o $\mathbb{R}$ é contínua à direita em $\mathbb{R}$ se $\lim\limits_{\substack{y \to x \\ y > x}} F(y) = F(x), \quad \forall x \in \mathbb{R}$.

O **extremo esquerdo do suporte** e o **extremo direito do suporte** da f.d. F estão definidos, respectivamente, por

$$x_{\inf} = \inf\big\{x \in \mathbb{R}:\ F(x) > 0\big\} \quad \text{e} \quad x_{\sup} = \sup\big\{x \in \mathbb{R}:\ F(x) < 1\big\}.$$

Definição 1.1.9 Duas funções de distribuição F e G dizem-se de **cauda equivalente** se o extremo direito do suporte de ambas for o mesmo e

$$\lim_{x \to x_{\sup}} \frac{\overline{F}(x)}{\overline{G}(x)} = c$$

para alguma constante $0 < c < \infty$.

Na realidade, as funções de distribuição são artefactos matemáticos com propriedades que são independentes de qualquer conjectura probabilística. No entanto, uma v.a. X definida num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ induz um outro espaço de probabilidade, a saber, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_{\mathrm{X}})$ por meio da correspondência,

$$\mathbb{P}_{\mathrm{X}}(B) = \mathbb{P}\big[\mathrm{X}^{-1}(B)\big] = \mathbb{P}\big\{\omega:\ \mathrm{X}(\omega) \in B\big\}, \quad \forall B \in \mathcal{B}(\mathbb{R}) \tag{1.1.1}$$

(ver [81]). Alguns autores (ver [12]) apelidam a medida de probabilidade $\mathbb{P}_{\mathrm{X}}$ definida em (1.1.1) de **distribuição** da variável aleatória X. Esta sugestiva designação acontece, justamente, porque $\mathbb{P}_{\mathrm{X}}$ está estreitamente relacionada com a classe das funções de distribuição que acabamos de apresentar. Com efeito, existe uma correspondência entre as funções de conjunto (1.1.1) e as funções de distribuição, que passamos a descrever no resultado abaixo (ver [81], página 46).

Teorema 1.1.2 *Dada uma medida de probabilidade* $\mathbb{P}$ *sobre* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *existe uma função de distribuição* F *tal que,*

$$\mathbb{P}\big(]-\infty, x]\big) = F(x), \quad \forall x \in \mathbb{R}. \tag{1.1.2}$$

Reciprocamente, dada uma função de distribuição F *existe uma única medida de probabilidade* $\mathbb{P}$ *definida em* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *que satisfaz* (1.1.2).

Demonstração. Consultar [19], páginas 25 e 27. ∎

O interesse do Teorema 1.1.2 reside também no facto de se poder evitar as funções de conjunto (em particular, a medida de probabilidade $\mathbb{P}_{\mathrm{X}}$) que, como é sabido, são sempre funções de difícil manuseamento. É então com naturalidade que surge o conceito de função de distribuição de uma variável aleatória.

Definição 1.1.10 Seja X uma v.a. definida num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Chamamos **função de distribuição** da v.a. X à função real de variável real F_{X} dada por,

$$F_{\mathrm{X}}(x) = \mathbb{P}_{\mathrm{X}}\big(]-\infty, x]\big) = \mathbb{P}\big\{\omega:\ \mathrm{X}(\omega) \leqslant x\big\}, \quad \forall x \in \mathbb{R}.$$

Observação 1.1.1 Note-se que dada uma função de distribuição F existe, em algum espaço de probabilidade, uma variável aleatória X para a qual $F(x) = \mathbb{P}\{X \leqslant x\}$.

Apresentamos agora dois tipos clássicos de variáveis aleatórias.

Exemplo 1.1.6

Uma v.a. X definida sobre $(\Omega, \mathcal{F}, \mathbb{P})$ diz-se **discreta** se existe um conjunto finito ou enumerável $E \subset \mathbb{R}$ tal que $\mathbb{P}\big\{X \in E\big\} = 1$. Em especial, se $E = \{x_1, x_2, \ldots\} \subset \mathbb{R}$ então a colecção de números não-negativos $\{p_i\}$ que satisfaz $p_i = \mathbb{P}\{X = x_i\}$ para todo o $i = 1, 2, \ldots$ e $\sum_{i=1}^{\infty} p_i = 1$ é chamada **função massa de probabilidade** da v.a. X.

Exemplo 1.1.7

Se X é uma v.a. definida sobre $(\Omega, \mathcal{F}, \mathbb{P})$ com função de distribuição F então X diz-se **absolutamente contínua** se existe uma função boreliana $f\colon \mathbb{R} \to \mathbb{R}$ não-negativa[4] tal que para qualquer número real x se tenha,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

A função f é chamada **função densidade de probabilidade** da v.a. X.

Uma função de distribuição G será chamada **absolutamente contínua** se existir uma função boreliana $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ tal que $G(x) = \int_{-\infty}^{x} g(t)\, dt$, $x \in \mathbb{R}$. A função g associada será apelidada sugestivamente de **densidade** já que necessariamente $\int_{-\infty}^{\infty} g(t)\, dt = 1$ e $g \geqslant 0$ a.e. (isto é, g será não-negativa em todo o $\mathbb{R}$ excepto num subconjunto de $\mathbb{R}$ que tenha medida de Lebesgue nula).

Sejam X e Y duas v.a.'s definidas num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. A v.a. X é **degenerada** num ponto c se $\mathbb{P}\{X = c\} = 1$ e $\mathbb{P}\{X \neq c\} = 0$, ou seja, se a f.d. de X for degenerada tendo c como único ponto de crescimento. Por outro lado, dizemos que X é **simétrica** relativamente a um ponto a se,

$$\mathbb{P}\big\{X \geqslant a + x\big\} = \mathbb{P}\big\{X \leqslant a - x\big\}, \quad \forall x \in \mathbb{R}$$

ou considerando a f.d. F_X da v.a. X tem-se, de um modo equivalente,

$$F_X(a - x) = 1 - F(a + x) + \mathbb{P}\big\{X = a + x\big\}, \quad \forall x \in \mathbb{R}.$$

O ponto a designa-se geralmente por **centro de simetria**. Se as v.a.'s X e Y tiverem a mesma função de distribuição i.e.

$$F_X(x) = F_Y(x), \quad \forall x \in \mathbb{R}$$

[4]Em boa verdade, $f(x)$ apenas terá de ser não-negativa *almost everywhere* (abreviadamente *a.e*), ou seja, terá de ser não-negativa em todo o $\mathbb{R}$ excepto num subconjunto de $\mathbb{R}$ que tenha medida de Lebesgue nula (ver Exemplo 1.1.4). Um conjunto $U \subset \mathbb{R}$ tem medida de Lebesgue nula se para todo o $\varepsilon > 0$, existir uma família numerável de intervalos de comprimento total menor que ε cuja união inclua U. Convém observar ainda que F, sendo o integral indefinido de f, é contínua. Tecnicamente, este integral é o de Lebesgue tendo-se consequentemente $f(x) = F'(x)$ a.e. (i.e. em todo o ponto, excepto num conjunto de medida de Lebesgue nula).

então dizemos que X e Y são **identicamente distribuídas** e indicamos $\mathrm{X} \stackrel{\mathrm{d}}{=} \mathrm{Y}$. Mais geralmente, dizemos ainda que X e Y (ou as correspondentes funções de distribuição) são do **mesmo tipo** se existem constantes $a > 0$ e $b \in \mathbb{R}$ tais que $\mathrm{X} \stackrel{\mathrm{d}}{=} a\,\mathrm{Y} + b$.

Em muitas experiências, uma observação é exprimida, não por uma única quantidade numérica, mas por várias quantidades numéricas separadas. Para pudermos descrever matematicamente tais experiências, torna-se indispensável a introdução do conceito de vector aleatório.

Definição 1.1.11 Dado um espaço amostral $(\Omega, \mathcal{F})$ chamamos **vector aleatório** a uma função $\mathbf{X}\colon \Omega \to \mathbb{R}^n$ cuja imagem inversa de qualquer boreliano de $\mathbb{R}^n$ é um acontecimento,

$$\mathbf{X}^{-1}(B) = \big\{\omega\colon\ \mathbf{X}(\omega) \in B\big\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

Observemos que se $(\Omega, \mathcal{F})$ for um espaço amostral e $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)\colon \Omega \to \mathbb{R}^n$ uma função então $\mathbf{X}$ é um vector aleatório se e só se cada X_i for uma variável aleatória.

À semelhança do caso unidimensional, também os vectores aleatórios estão estreitamente ligados a uma classe de funções apropriadamente definidas: a classe das funções de distribuição n-dimensionais.

Definição 1.1.12 Uma função F definida em todo o $\mathbb{R}^n$ com valores em $\mathbb{R}$ que satisfaça,

(D1) $\displaystyle\lim_{x_i \to -\infty} F(x_1, \ldots, x_n) = F(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_n) = 0, \quad i = 1, \ldots, n.$

(D2) $\displaystyle\lim_{(x_1,\ldots,x_n)\to(\infty,\ldots,\infty)} F(x_1, \ldots, x_n) = F(\infty, \ldots, \infty) = 1.$

(D3) $\displaystyle\lim_{\substack{(y_1,\ldots,y_n)\to(x_1,\ldots,x_n)\\ y_i > x_i,\, \forall i}} F(y_1, \ldots, y_n) = F(x_1, \ldots, x_n).$

(D4)

$$F(y_1, \ldots, y_n) - \sum_{i=1}^{n} F(y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_n) +$$
$$+ \sum_{i \leqslant i < j \leqslant n} F(y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_{j-1}, x_j, y_{j+1}, \ldots, y_n) + \ldots + (-1)^n F(x_1, \ldots, x_n) \geqslant 0$$

sempre que $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathbb{R}^n$ com $-\infty < x_i \leqslant y_i < \infty$, $i = 1, \ldots, n$.

é chamada **função de distribuição n-dimensional**.

Com efeito, um vector aleatório $\mathbf{X}$ definido num espaço de probabilidade $(\Omega,\ \mathcal{F}, \mathbb{P})$ ainda induz o espaço de probabilidade $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{\mathbf{X}})$ por meio da correspondência,

$$\mathbb{P}_{\mathbf{X}}(B) = \mathbb{P}\big[\mathbf{X}^{-1}(B)\big] = \mathbb{P}\big\{\omega\colon\ \mathbf{X}(\omega) \in B\big\}, \quad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

É habitual chamar **distribuição n-dimensional** à medida de probabilidade $\mathbb{P}_{\mathbf{X}}$ pois, à semelhança do caso unidimensional, $\mathbb{P}_{\mathbf{X}}$ está intimamente relacionada com a classe de funções de distribuição

n-dimensionais. Efectivamente, o Teorema 1.1.2 admite uma versão n-dimensional que passamos a anunciar.

Teorema 1.1.3 *Dada uma medida de probabilidade $\mathbb{P}$ sobre $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ existe uma função de distribuição n-dimensional F tal que,*

$$\mathbb{P}\big(]-\infty, x_1] \times \ldots \times]-\infty, x_n]\big) = F(x_1, \ldots, x_n), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n. \tag{1.1.3}$$

Reciprocamente, dada uma função de distribuição n-dimensional F existe uma única medida de probabilidade $\mathbb{P}$ definida em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ que satisfaz (1.1.3).

Demonstração. Seja $\mathbb{P}$ uma medida de probabilidade sobre $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ e consideremos a função $F\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ definida por $F(x_1, \ldots, x_n) = \mathbb{P}\big(]-\infty, x_1] \times \ldots \times]-\infty, x_n]\big)$. Supondo os rectângulos de $\mathbb{R}^n$,

$$S_{x_1,\ldots,x_n} =]-\infty, x_1] \times \ldots \times]-\infty, x_n]$$

temos:[5]

(D1) Se para algum $i = 1, \ldots, n$, $x_i \downarrow -\infty$ então,

$$S_{x_1,\ldots,x_n} \downarrow\,]-\infty, x_1] \times \ldots \times]-\infty, x_{i-1}] \times \emptyset \times]-\infty, x_{i+1}] \times \ldots \times]-\infty, x_n] = \emptyset$$

e o **Teorema 10.2** de [12] na página 162 permite escrever,

$$\lim_{x_i \downarrow -\infty} F(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) = \lim_{x_i \downarrow -\infty} \mathbb{P}\big(S_{x_1,\ldots,x_n}\big) = \mathbb{P}(\emptyset) = 0.$$

(D2) Uma vez que $S_{x_1,\ldots,x_n} \uparrow \mathbb{R}^n$ quando $(x_1, \ldots, x_n) \uparrow (\infty, \ldots, \infty)$ temos do **Teorema 10.2** de [12],

$$\lim_{(x_1,\ldots,x_n)\uparrow(\infty,\ldots,\infty)} F(x_1, \ldots, x_n) = \lim_{(x_1,\ldots,x_n)\uparrow(\infty,\ldots,\infty)} \mathbb{P}\big(S_{x_1,\ldots,x_n}\big) = \mathbb{P}(\mathbb{R}^n) = 1.$$

(D3) Se $(y_1, \ldots, y_n) \downarrow (x_1, \ldots, x_n)$ então $S_{y_1,\ldots,y_n} \downarrow S_{x_1,\ldots,x_n}$ obtendo-se, novamente do **Teorema 10.2** de [12],

$$\lim_{(y_1,\ldots,y_n)\downarrow(x_1,\ldots,x_n)} F(y_1, \ldots, y_n) = \lim_{(y_1,\ldots,y_n)\downarrow(x_1,\ldots,x_n)} \mathbb{P}\big(S_{y_1,\ldots,y_n}\big) = \mathbb{P}\big(S_{x_1,\ldots,x_n}\big) = F(x_1, \ldots, x_n)$$

o que estabelece a continuidade à direita de F.

[5] No que se segue adoptaremos as seguintes notações:

(i) $y \uparrow x$ significará que $y < x$ e $y \to x$; $y \downarrow x$ significará que $y > x$ e $y \to x$.

(ii) $(y_1, \ldots, y_n) \uparrow (x_1, \ldots, x_n)$ significará que $y_i \uparrow x_i$ para cada $i = 1, \ldots, n$; $(y_1, \ldots, y_n) \downarrow (x_1, \ldots, x_n)$ significará que $y_i \downarrow x_i$ para cada $i = 1, \ldots, n$.

(iii) $A_m \uparrow A$ significará que $A_1 \subset A_2 \subset \ldots$ e $A = \bigcup_{m=1}^{\infty} A_m$; $A_m \downarrow A$ significará que $A_1 \supset A_2 \supset \ldots$ e $A = \bigcap_{m=1}^{\infty} A_m$.

(D4) Observando que qualquer rectângulo de $\mathbb{R}^n$ limitado $A =]a_1, b_1] \times \ldots \times]a_n, b_n]$ pode ser escrito como $A = S_{b_1,\ldots,b_n} \setminus \big(S_{a_1,b_2,\ldots,b_n} \cup S_{b_1,a_2,\ldots,b_n} \cup \ldots S_{b_1,b_2,\ldots,a_n}\big)$ tem-se da fórmula de inclusão-exclusão (ver [12], página 24),

$$\begin{aligned}
0 \leqslant \mathbb{P}(A) &= \mathbb{P}(S_{b_1,\ldots,b_n}) - \mathbb{P}\big(S_{a_1,b_2,\ldots,b_n} \cup S_{b_1,a_2,\ldots,b_n} \cup \ldots S_{b_1,b_2,\ldots,a_n}\big) \\
&= F(b_1,\ldots,b_n) - \sum_{i=1}^{n} F(b_1,\ldots,b_{i-1},a_i,b_{i+1},\ldots,b_n) + \\
&+ \sum_{1\leqslant i<j\leqslant n} F(b_1,\ldots,b_{i-1},a_i,b_{i+1},\ldots,b_{j-1},a_j,b_{j+1},\ldots,b_n) + \ldots + (-1)^n F(a_1,\ldots,a_n)
\end{aligned}$$

para quaisquer $(a_1,\ldots,a_n), (b_1,\ldots,b_n) \in \mathbb{R}^n$ com $-\infty < a_i \leqslant b_i < \infty$, $i = 1,\ldots,n$.

Reciprocamente, se F for uma função de distribuição n-dimensional então as condições (D3) e (D4) garantem que existe uma única medida $\mathbb{P}$ sobre a σ-álgebra $\mathcal{B}(\mathbb{R}^n)$ tal que

$$\begin{aligned}
\mathbb{P}\big(]a_1,b_1]\times\ldots\times]a_n,b_n]\big) &= F(b_1,\ldots,b_n) - \sum_{i=1}^{n} F(b_1,\ldots,b_{i-1},a_i,b_{i+1},\ldots,b_n) + \\
&+ \sum_{1\leqslant i<j\leqslant n} F(b_1,\ldots,b_{i-1},a_i,b_{i+1},\ldots,b_{j-1},a_j,b_{j+1},\ldots,b_n) + \ldots + (-1)^n F(a_1,\ldots,a_n)
\end{aligned}$$

(ver **Teorema 12.5** de [12], página 177). Fazendo $a_1 \to -\infty$ vem,

$$]a_1,b_1]\times]a_2,b_2] \times \ldots \times]a_n,b_n] \uparrow]-\infty,b_1]\times]a_2,b_2] \times \ldots \times]a_n,b_n]$$

donde

$$\begin{aligned}
\mathbb{P}\big(]-\infty,b_1]\times]a_2,b_2] \times \ldots \times]a_n,b_n]\big) &= \lim_{a_1\to-\infty} \mathbb{P}\big(]a_1,b_1]\times]a_2,b_2] \times \ldots \times]a_n,b_n]\big) \\
&= F(b_1,\ldots,b_n) - \sum_{i=2}^{n} F(b_1,\ldots,b_{i-1},a_i,b_{i+1},\ldots,b_n) + \\
&+ \sum_{2\leqslant i<j\leqslant n} F(b_1,\ldots,b_{i-1},a_i,b_{i+1},\ldots,b_{j-1},a_j,b_{j+1},\ldots,b_n) + \\
&\qquad + \ldots + (-1)^{n-1} F(b_1,a_2,\ldots,a_n)
\end{aligned}$$

pelo **Teorema 10.2** de [12]. De forma análoga, calculando sucessivamente os limites quando $a_2 \to -\infty$, ..., $a_n \to -\infty$ obtém-se $\mathbb{P}\big(]-\infty,b_1] \times \ldots \times]-\infty,b_n]\big) = F(b_1,\ldots,b_n)$. Visto que $]-\infty,b_1] \times \ldots \times]-\infty,b_n] \uparrow \mathbb{R}^n$ quando $(b_1,\ldots,b_n) \uparrow (\infty,\ldots,\infty)$ tem-se da condição (D2) e do **Teorema 10.2** de [12],

$$\mathbb{P}(\mathbb{R}^n) = \lim_{(b_1,\ldots,b_n)\uparrow(\infty,\ldots,\infty)} \mathbb{P}\big(]-\infty,b_1] \times \ldots \times]-\infty,b_n]\big) = \lim_{(b_1,\ldots,b_n)\uparrow(\infty,\ldots,\infty)} F(b_1,\ldots,b_n) = 1$$

pelo que $\mathbb{P}$ é uma medida de probabilidade que satisfaz (1.1.3). ∎

Estamos agora em condições de generalizar a Definição 1.1.10.

Definição 1.1.13 Se $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ for um vector aleatório definido no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ chamamos **função de distribuição conjunta** de $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ à função $F_{\mathrm{X}_1,\ldots,\mathrm{X}_n}$ definida em todo o $\mathbb{R}^n$ com valores reais dada por,

$$F_{\mathrm{X}_1,\ldots,\mathrm{X}_n}(x_1, \ldots, x_n) = \mathbb{P}\big\{\mathrm{X}_1 \leqslant x_1, \ldots, \mathrm{X}_n \leqslant x_n\big\}, \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Observação 1.1.2 Convém sublinhar que dada uma função de distribuição conjunta F existe sempre, nalgum espaço de probabilidade, um vector aleatório $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ para o qual $F(x_1, \ldots, x_n) = \mathbb{P}\big\{\mathrm{X}_1 \leqslant x_1, \ldots, \mathrm{X}_n \leqslant x_n\big\}$.

Exemplo 1.1.8

À semelhança do Exemplo 1.1.7 um vector aleatório $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ definido em $(\Omega, \mathcal{F}, \mathbb{P})$ diz-se **absolutamente contínuo** se existe uma função boreliana[6] não-negativa $f\colon \mathbb{R}^n \to \mathbb{R}$ tal que para cada $(x_1, \ldots, x_n) \in \mathbb{R}^n$ se tenha,

$$F_{\mathrm{X}_1,\ldots,\mathrm{X}_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n)\, dt_1 \ldots dt_n$$

onde $F_{\mathrm{X}_1,\ldots,\mathrm{X}_n}$ é a função de distribuição conjunta de $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$. A função f é chamada **função de densidade conjunta** do vector $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$.

Um caso especial de uma função de densidade conjunta, com grande interesse neste trabalho, é apresentado de seguida.

Definição 1.1.14 Dizemos que um vector aleatório $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ absolutamente contínuo definido num espaço de probabilidade tem **simetria radial** se a sua função de densidade conjunta for,

$$f_{\mathrm{X}_1,\ldots,\mathrm{X}_n}(x_1, \ldots, x_n) = g(r), \qquad r = \sqrt{x_1^2 + \ldots + x_n^2}$$

dependendo apenas da distância à origem através de alguma função não-negativa g.

A independência pode ser vista como o único e mais importante conceito em teoria de probabilidades, demarcando-se da teoria da medida e alimentando um desenvolvimento à parte. No curso desta evolução, a teoria de probabilidades tem sido fortalecida pelas suas ligações ao mundo real e, de facto, a definição de independência constitui uma abstracção de uma noção altamente intuitiva e empírica.

Definição 1.1.15 Os acontecimentos $A_1, \ldots, A_n$ $(n \geqslant 2)$ definidos num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ dizem-se (mutuamente) **independentes** se

$$\mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_m}\right) = \mathbb{P}\left(A_{i_1}\right) \ldots \mathbb{P}\left(A_{i_m}\right)$$

para todo o $1 \leqslant i_1 < \ldots < i_m \leqslant n$ e para todo o $m = 2, \ldots, n$ (i.e. se todas as combinações satisfazem a regra do produto).

[6]Uma função $f\colon \mathbb{R}^n \to \mathbb{R}$ é boreliana se for $\mathcal{B}(\mathbb{R}^n)$-mensurável considerando $(\Omega, \mathcal{F}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ e $(\Sigma, \mathcal{G}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Observação 1.1.3 Os acontecimentos $A_1, A_2, \ldots$ são independentes se para todo o $n \geqslant 2$, $A_1, \ldots, A_n$ são independentes.

Estamos agora em condições de anunciar a segunda parte do Teorema 1.1.1 conhecida por segundo lema de Borel-Cantelli.

Teorema 1.1.4 (segundo lema de Borel-Cantelli) *Seja* $\{A_n,\, n \geqslant 1\}$ *uma sucessão de acontecimentos independentes num espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbb{P})$ *então,*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \quad \Longrightarrow \quad \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

Demonstração. Consultar [91], página 40. ∎

Sejam $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ $(n \geqslant 2)$ v.a's definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e consideremos o vector aleatório $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ também definido em $(\Omega, \mathcal{F}, \mathbb{P})$. Informalmente, as v.a.'s X_i são independentes se e só se quaisquer acontecimentos determinados por qualquer grupo de v.a.'s distintas são independentes.

Definição 1.1.16 As v.a.'s $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ são (mutuamente) **independentes** se,

$$\mathbb{P}\{\mathrm{X}_1 \in B_1, \mathrm{X}_2 \in B_2, \ldots, \mathrm{X}_n \in B_n\} = \mathbb{P}\{\mathrm{X}_1 \in B_1\}\, \mathbb{P}\{\mathrm{X}_2 \in B_2\} \ldots \mathbb{P}\{\mathrm{X}_n \in B_n\}$$

para todos os subconjuntos $B_1, B_2, \ldots, B_n \in \mathcal{B}(\mathbb{R})$.

Ao longo de todo este texto, usaremos a abreviatura i.i.d. para indicar que v.a.'s $\mathrm{X}_1, \ldots, \mathrm{X}_n$ definidas num espaço de probabilidade são (mutuamente) independentes e identicamente distribuídas. Decorre ainda que as v.a.'s $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ são independentes se a função de distribuição do vector aleatório $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ puder ser factorizada como o produto das funções de distribuição individuais de cada X_i.

Teorema 1.1.5 *Sejam* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *v.a.'s definidas em* $(\Omega, \mathcal{F}, \mathbb{P})$*, respectivamente, com funções de distribuição* $F_{\mathrm{X}_1}, \ldots, F_{\mathrm{X}_n}$ *e* $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ *o vector aleatório com função de distribuição conjunta* $F_{\mathrm{X}_1,\ldots,\mathrm{X}_n}$.

(a) *Se* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *são independentes então,*

$$F_{\mathrm{X}_1,\ldots,\mathrm{X}_n}(x_1, \ldots, x_n) = F_{\mathrm{X}_1}(x_i) \ldots F_{\mathrm{X}_n}(x_n), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

(b) *Se existem funções* $F_1, \ldots, F_n$ *tais que* $\lim_{x\to\infty} F_i(x) = 1$ *para todo o* $i = 1, \ldots, n$ *e,*

$$F_{\mathrm{X}_1,\ldots,\mathrm{X}_n}(x_1, \ldots, x_n) = F_1(x_1) \ldots F_n(x_n), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

então $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *são independentes e* $F_i = F_{\mathrm{X}_i}$ *para todo o* $i = 1, \ldots, n$.

Demonstração. Consultar [42], página 61. ∎

Corolário 1.1.1

(a) *Se* $X_1, \ldots, X_n$ *são v.a.'s independentes, absolutamente contínuas e possuem densidades* $f_{X_1}, \ldots,$ f_{X_n} *então a função,*

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n$$

é a função de densidade conjunta do vector aleatório $(X_1, \ldots, X_n)$.

(b) *Reciprocamente, se o vector* $(X_1, \ldots, X_n)$ *é absolutamente contínuo e tem função de densidade conjunta* f *satisfazendo,*

$$f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

onde $f_i \geqslant 0$ *e* $\int_{-\infty}^{\infty} f_i(t)\, dt = 1$, $\forall i$ *então as v.a.'s* $X_1, \ldots, X_n$ *são independentes e* f_i *é a densidade de* X_i *para todo o* $i = 1, \ldots, n$.

Demonstração. Consultar [42], página 63. ∎

Apresentamos, em seguida, duas desigualdades relevantes para somas de v.a.'s independentes conhecidas por desigualdades de tipo Kolmogorov (ver [49], página 44).

Teorema 1.1.6 (desigualdade de Lévy) *Se* $X_1, \ldots, X_n$ *são v.a.'s independentes e simétricas relativamente à origem então para todo o* $\varepsilon > 0$,

$$\mathbb{P}\left\{\max_{1 \leqslant j \leqslant n} \left|\sum_{i=1}^{j} X_i\right| \geqslant \varepsilon\right\} \leqslant 2\,\mathbb{P}\left\{\left|\sum_{i=1}^{n} X_i\right| \geqslant \varepsilon\right\}.$$

Demonstração. Consultar [18], página 72. ∎

Teorema 1.1.7 *Se* $X_1, \ldots, X_n$ *são v.a.'s independentes e para algum* $\delta > 0$ *e* $\alpha < 1$,

$$\mathbb{P}\left\{\left|\sum_{i=j+1}^{n} X_i\right| \geqslant \delta\right\} \leqslant \alpha, \quad j = 0, 1, \ldots, n-1$$

então para todo o $\varepsilon > 0$,

$$\mathbb{P}\left\{\max_{1 \leqslant j \leqslant n} \left|\sum_{i=1}^{j} X_i\right| \geqslant \delta + \varepsilon\right\} \leqslant \frac{1}{1-\alpha}\,\mathbb{P}\left\{\left|\sum_{i=1}^{n} X_i\right| \geqslant \varepsilon\right\}.$$

Demonstração. A demonstração é uma simples adaptação da desigualdade de Ottaviani apresentada em [24] ou [92]. ∎

Antes de prosseguirmos, façamos a extensão das operações usuais de soma e multiplicação de números a $[0, \infty]$ convencionando que: $0 \cdot \infty = \infty \cdot 0 = 0$ e $x \cdot \infty = \infty \cdot x = \infty$ quando $0 < x \leqslant \infty$.

Se x for um real finito convencionaremos ainda que $\infty - x = \infty$ e $x - \infty = -\infty$. Em teoria da medida, sendo $\mathfrak{f}$ uma função mensurável não-negativa no espaço de medida $(\Omega, \mathcal{F}, \mathfrak{m})$ e $\{A_i\}$ uma decomposição finita de Ω em conjuntos $\mathcal{F}$-mensuráveis define-se **integral** de $\mathfrak{f}$ por,

$$\int_\Omega \mathfrak{f}(\omega)\, \mathfrak{m}(d\omega) = \sup \sum_i \left[\inf_{\omega \in A_i} \mathfrak{f}(\omega) \right] \mathfrak{m}(A_i)$$

onde o supremo é tomado sobre todas as decomposições finitas $\{A_i\}$ de Ω em conjuntos $\mathcal{F}$-mensuráveis.[7] Dada uma função geral $\mathfrak{f}$, consideremos a sua parte positiva $\mathfrak{f}^+ = \max(\mathfrak{f}, 0)$ e a sua parte negativa $\mathfrak{f}^- = -\min(\mathfrak{f}, 0)$. Visto que $\mathfrak{f}^+ \geqslant 0$, $\mathfrak{f}^- \geqslant 0$ e $\mathfrak{f} = \mathfrak{f}^+ - \mathfrak{f}^-$ podemos definir mais geralmente,

$$\int_\Omega \mathfrak{f}(\omega)\, \mathfrak{m}(d\omega) = \int_\Omega \mathfrak{f}^+(\omega)\, \mathfrak{m}(d\omega) - \int_\Omega \mathfrak{f}^-(\omega)\, \mathfrak{m}(d\omega)$$

excepto quando $\int_\Omega \mathfrak{f}^+(\omega)\, \mathfrak{m}(d\omega) = \int_\Omega \mathfrak{f}^-(\omega)\, \mathfrak{m}(d\omega) = \infty$, situação em que $\mathfrak{f}$ não tem integral. Se $\int_\Omega \mathfrak{f}^+(\omega)\, \mathfrak{m}(d\omega)$ e $\int_\Omega \mathfrak{f}^-(\omega)\, \mathfrak{m}(d\omega)$ forem finitos então $\mathfrak{f}$ diz-se **integrável** e $\int_\Omega \mathfrak{f}(\omega)\, \mathfrak{m}(d\omega)$ é o seu integral definido. Se $\int_\Omega \mathfrak{f}^+(\omega)\, \mathfrak{m}(d\omega) = \infty$ e $\int_\Omega \mathfrak{f}^-(\omega)\, \mathfrak{m}(d\omega) < \infty$ (resp. $\int_\Omega \mathfrak{f}^+(\omega)\, \mathfrak{m}(d\omega) < \infty$ e $\int_\Omega \mathfrak{f}^-(\omega)\, \mathfrak{m}(d\omega) = \infty$) então $\mathfrak{f}$ não é integrável mas o valor do integral definido $\int_\Omega \mathfrak{f}(\omega)\, \mathfrak{m}(d\omega)$ é ∞ (resp. $-\infty$). O integral de $\mathfrak{f}$ sobre um elemento A de $\mathcal{F}$ é definido por,

$$\int_A \mathfrak{f}(\omega) \mathfrak{m}(d\omega) = \int_\Omega I_A(\omega) \mathfrak{f}(\omega) \mathfrak{m}(d\omega)$$

onde I_A é o **indicador** de A, ou seja, a função $I_A \colon \Omega \to \{0, 1\}$ definida por $I_A(\omega) = 1$ se $\omega \in A$ e $I_A(\omega) = 0$ se $\omega \notin A$.

O leitor tem certamente uma ideia intuitiva de média de uma determinada quantidade aleatória. Frases como 'este carro gasta uma média de 8 litros aos 100Km' ou 'o vencedor da décima-sexta etapa do Tour de France 2004 (contra-relógio individual Bourg D'Oisans - L'Alpe D'Huez) fez uma média de 23.435Km/h' são-nos bastante familiares. Formalmente, estamos em condições de estabelecer a noção de valor médio ou valor esperado de uma variável aleatória.

Definição 1.1.17 O **valor médio** de uma v.a. X definida no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ é o integral de X relativamente à medida de probabilidade $\mathbb{P}$,

$$\mathbb{E}(\mathrm{X}) = \int_\Omega \mathrm{X}(\omega)\, \mathbb{P}(d\omega).$$

Em analogia com o exposto anteriormente, dada uma v.a. X não-negativa o seu valor médio $\mathbb{E}(\mathrm{X})$ existe sempre (podendo ser infinito). O valor médio de uma qualquer v.a. X está definido se pelo menos um dos valores médios $\mathbb{E}(\mathrm{X}^+)$ ou $\mathbb{E}(\mathrm{X}^-)$ for finito tendo-se $\mathbb{E}(\mathrm{X}) = \mathbb{E}(\mathrm{X}^+) - \mathbb{E}(\mathrm{X}^-)$. Enfim,

[7] Note-se que definições alternativas de integral podem ser esboçadas como, por exemplo, a aproximação de Darboux-Young: indicando o **integral inferior** e **superior**, respectivamente, por $\underline{\int}_\Omega \mathfrak{f}(\omega)\, \mathfrak{m}(d\omega) = \sup \sum_i \left[\inf_{\omega \in A_i} \mathfrak{f}(\omega)\right] \mathfrak{m}(A_i)$ e $\overline{\int}_\Omega \mathfrak{f}(\omega)\, \mathfrak{m}(d\omega) = \inf \sum_i \left[\sup_{\omega \in A_i} \mathfrak{f}(\omega)\right] \mathfrak{m}(A_i)$ onde o ínfimo e o supremo são tomados sobre todas as decomposições finitas $\{A_i\}$ de Ω em conjuntos $\mathcal{F}$-mensuráveis, podemos definir integral como o valor comum $\underline{\int}_\Omega \mathfrak{f}(\omega)\, \mathfrak{m}(d\omega) = \overline{\int}_\Omega \mathfrak{f}(\omega)\, \mathfrak{m}(d\omega)$ (ver [12] para mais detalhes).

X é integrável se $\mathbb{E}\,|\mathrm{X}| < \infty$. O integral $\int_A \mathrm{X}(\omega)\mathbb{P}(d\omega)$ sobre o conjunto A é definido por $\mathbb{E}(I_A\mathrm{X})$ onde I_A é o indicador de A.

Sendo $p > 0$ e X uma v.a. num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ chamamos **momento absoluto de ordem p** de X a $\mathbb{E}\big[\,|\mathrm{X}|^p\,\big]$. Em particular, se X for integrável chamamos **variância** de X ao valor $\mathbb{V}(\mathrm{X}) = \mathbb{E}\left[\big(\mathrm{X} - \mathbb{E}(\mathrm{X})\big)^2\right]$. Associado a cada espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ estão os **espaços $\mathscr{L}_p$** de todas as v.a.'s X com momento absoluto de ordem p finito,

$$\mathscr{L}_p(\Omega, \mathcal{F}, \mathbb{P}) = \Big\{\mathrm{X}\colon \Omega \to \mathbb{R} \;\Big|\; \mathrm{X} \text{ é variável aleatória} \quad \text{e} \quad \mathbb{E}\big[\,|\mathrm{X}|^p\,\big] < \infty\Big\}.$$

Quando $p \geqslant 1$ podemos definir sobre o espaço vectorial $\mathscr{L}_p(\Omega, \mathcal{F}, \mathbb{P})$ a norma $\mathbb{E}^{1/p}\big[\,|\mathrm{X}|^p\,\big]$ obtendo-se, assim, um espaço vectorial normado (mais geralmente $\mathscr{L}_p(\Omega, \mathcal{F}, \mathbb{P})$ é um espaço de Banach real).[8]

A integração num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ constitui uma teoria excepcional, mas para efectuarmos cálculos temos de mudar para um outro espaço, na circunstância $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$.

Teorema 1.1.8 (fórmula de mudança de variável) *Seja* X *uma variável aleatória de* $(\Omega, \mathcal{F})$ *e* $\mathbb{P}_\mathrm{X}$ *a respectiva distribuição definida por* (1.1.1). *Se* $h\colon \mathbb{R} \longrightarrow \mathbb{R}$ *é uma função boreliana e* $\mathbb{E}\big[\,|h(\mathrm{X})|\,\big] < \infty$ *então,*

$$\mathbb{E}\big[h(\mathrm{X})\big] = \int_{-\infty}^{\infty} h(y)\,\mathbb{P}_\mathrm{X}(dy). \tag{1.1.4}$$

Demonstração. Consultar [25], página 18. ■

Observação 1.1.4 De notar que se um dos membros de (1.1.4) é infinito então o mesmo acontece ao outro membro.

Exemplo 1.1.9

Se X é uma v.a. discreta com função massa de probabilidade $p_i = \mathbb{P}\{\mathrm{X} = x_i\}$, $i = 1, 2, \ldots$ e $h\colon \mathbb{R} \longrightarrow \mathbb{R}$ é uma função boreliana então

$$\mathbb{E}\big[h(\mathrm{X})\big] = \sum_{i=1}^{\infty} h(x_i)p_i$$

se $\sum_{i=1}^{\infty} |h(x_i)|\, p_i < \infty$.

Exemplo 1.1.10

Se X é uma v.a. absolutamente contínua com função de densidade f e $h\colon \mathbb{R} \longrightarrow \mathbb{R}$ é uma função boreliana então

$$\mathbb{E}\big[h(\mathrm{X})\big] = \int_{-\infty}^{\infty} h(x)f(x)\,dx$$

se $\int_{-\infty}^{\infty} |h(x)|\, f(x)\,dx < \infty$.

A relação (1.1.4) pode ser extendida às variáveis aleatórias n-dimensionais e neste caso a mudança faz-se para o espaço $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_\mathbf{X})$.

[8] Mais geralmente, se $0 < p < 1$ então a aplicação $\mathsf{f}(\omega) \mapsto \big[\int_\Omega |\mathsf{f}(\omega)|^p\, \mathfrak{m}(d\omega)\big]^{\frac{1}{p}}$ definida sobre $\mathscr{L}_p(\Omega, \mathcal{F}, \mathfrak{m})$ não constitui uma norma (ver [39] para mais detalhes).

Teorema 1.1.9 *Seja* $\mathbf{X}$ *um vector aleatório definido em* $(\Omega, \mathcal{F}, \mathbb{P})$ *e* $\mathbb{P}_{\mathbf{X}}$ *a distribuição definida por* $\mathbb{P}_{\mathbf{X}}(B) = \mathbb{P}\big\{\omega\colon\ \mathbf{X}(\omega) \in B\big\}$, $\forall B \in \mathcal{B}(\mathbb{R}^n)$. *Se* $h\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ *é uma função* $\mathcal{B}(\mathbb{R}^n)$*-mensurável e* $\mathbb{E}\big[\,|h(\mathbf{X})|\,\big] < \infty$ *então,*

$$\mathbb{E}\big[h(\mathbf{X})\big] = \int_{\mathbb{R}^n} h(\boldsymbol{y})\,\mathbb{P}_{\mathbf{X}}(d\boldsymbol{y}).$$

Demonstração. Consultar [25], página 18. ∎

Definição 1.1.18 Dada uma v.a. X num espaço de probabilidade definimos **supremo quase certo** de X por,

$$\text{sup q.c. X} = \inf\big\{c\colon\ c \geqslant 0 \ \text{ e } \ \mathbb{P}\{\mathrm{X} > c\} = 0\big\}.$$

Anunciamos de seguida, sem demonstração, uma desigualdade fundamental em teoria de probabilidades que estabelece propriedades relativas ao comportamento de qualquer v.a. independentemente da sua distribuição.

Teorema 1.1.10 (desigualdade de Chebyshev) *Seja* X *uma v.a. e* $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ *uma função não-negativa, par*[9] *e não-decrescente em* $[0, \infty[$ *então para todo o* $\varepsilon > 0$,

$$\frac{\mathbb{E}\big[g(\mathrm{X})\big] - g(\varepsilon)}{\text{sup q.c. } g(\mathrm{X})} \leqslant \mathbb{P}\big\{\,|\mathrm{X}| \geqslant \varepsilon\big\} \leqslant \frac{\mathbb{E}\big[g(\mathrm{X})\big]}{g(\varepsilon)}.$$

Se g *for não-decrescente em* $\mathbb{R}$ *então para todo o* $\varepsilon > 0$,

$$\frac{\mathbb{E}\big[g(\mathrm{X})\big] - g(\varepsilon)}{\text{sup q.c. } g(\mathrm{X})} \leqslant \mathbb{P}\big\{\mathrm{X} \geqslant \varepsilon\big\} \leqslant \frac{\mathbb{E}\big[g(\mathrm{X})\big]}{g(\varepsilon)}.$$

Demonstração. Consultar [55], página 159. ∎

Observação 1.1.5 Em particular, fazendo $g(t) = |t|^r$, $r > 0$ no Teorema 1.1.10 obtém-se para todo o $\varepsilon > 0$,

$$\frac{\mathbb{E}\big[\,|\mathrm{X}|^r\,\big] - \varepsilon^r}{\text{sup q.c. } |\mathrm{X}|^r} \leqslant \mathbb{P}\big\{\,|\mathrm{X}| \geqslant \varepsilon\big\} \leqslant \frac{\mathbb{E}\big[\,|\mathrm{X}|^r\,\big]}{\varepsilon^r}.$$

i.e. a conhecida **desigualdade de Markov**.

O próximo resultado abre a porta para uma definição central na moderna teoria de probabilidades.

Teorema 1.1.11 *Seja* $(\Omega, \mathcal{F}, \mathbb{P})$ *um espaço de probabilidade e* X *uma v.a. com* $\mathbb{E}\,|\mathrm{X}| < \infty$. *Se* $\mathcal{H}$ *for uma* σ*-álgebra contida em* $\mathcal{F}$ *então existe uma v.a.* Y *tal que,*

(a) Y *é* $\mathcal{H}$*-mensurável.*

(b) $\mathbb{E}\,|\mathrm{Y}| < \infty$.

[9]Uma função $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ é par se $g(x) = g(-x)$, $\forall x \in \mathbb{R}$.

(c) *Para todo o conjunto H em $\mathcal{H}$ tem-se,*

$$\int_H \mathrm{Y}(\omega)\,\mathbb{P}(d\omega) = \int_H \mathrm{X}(\omega)\,\mathbb{P}(d\omega), \quad \forall H \in \mathcal{H}.$$

Além disso, se Y' *for outra v.a. com as propriedades* (a), (b) *e* (c) *então* $\mathrm{Y}' = \mathrm{Y}$ *q.c.* (*i.e.* $\mathbb{P}\{\mathrm{Y}' = \mathrm{Y}\} = 1$).

Demonstração. Consultar [91], página 84. ■

Definição 1.1.19 Seja X uma v.a. de $\mathscr{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ e $\mathcal{H}$ uma σ-álgebra contida em $\mathcal{F}$. Uma v.a. Y com as propriedades (a), (b) e (c) do Teorema 1.1.9 é chamada (uma versão da) **esperança condicional** de X dada a σ-álgebra $\mathcal{H}$ e é denotada escrevendo,

$$\mathrm{Y} = \mathbb{E}\{\mathrm{X}\,|\,\mathcal{H}\} \quad \text{q.c.}$$

Algumas propriedades importantes da esperança condicional são apresentadas em baixo.

Teorema 1.1.12 *Seja* X *uma v.a. de* $\mathscr{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ *e* $\mathcal{H}$ *uma* σ*-álgebra contida em* $\mathcal{F}$*. Então,*

(a) *se* $\mathrm{Y} \in \mathscr{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ *tem-se*

$$\mathbb{E}\{a\mathrm{X} + b\mathrm{Y}\,|\,\mathcal{H}\} = a\,\mathbb{E}\{\mathrm{X}\,|\,\mathcal{H}\} + b\,\mathbb{E}\{\mathrm{Y}\,|\,\mathcal{H}\} \quad q.c. \qquad (a, b \in \mathbb{R}).$$

(b) $\mathbb{E}\big\{\mathbb{E}\{\mathrm{X}\,|\,\mathcal{H}\}\big\} = \mathbb{E}(\mathrm{X})$.

(c) *se* $\mathcal{G}$ *é uma* σ*-álgebra contida em* $\mathcal{H}$ *tem-se*

$$\mathbb{E}\big\{\mathbb{E}\{\mathrm{X}\,|\,\mathcal{H}\}\,|\,\mathcal{G}\big\} = \mathbb{E}\{\mathrm{X}\,|\,\mathcal{G}\} = \mathbb{E}\big\{\mathbb{E}\{\mathrm{X}\,|\,\mathcal{G}\}\,|\,\mathcal{H}\big\} \quad q.c.$$

(d) *se* Y *é* $\mathcal{H}$*-mensurável e* $\mathrm{XY} \in \mathscr{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ *tem-se*

$$\mathbb{E}\{\mathrm{XY}\,|\,\mathcal{H}\} = \mathrm{Y}\,\mathbb{E}\{\mathrm{X}\,|\,\mathcal{H}\} \quad q.c.^{10}$$

Demonstração. Consultar [85], página 215. ■

Se X for uma v.a. num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ tal que $\mathbb{E}\,|\mathrm{X}| < \infty$ e $\{\mathrm{Y}_j,\, j \in J\}$ uma família de v.a.'s de $(\Omega, \mathcal{F}, \mathbb{P})$ então denotamos,

$$\mathbb{E}(\mathrm{X}\,|\,\mathrm{Y}_j,\, j \in J) = \mathbb{E}\big\{\mathrm{X}\,|\,\sigma(\mathrm{Y}_j,\, j \in J)\big\}$$

[10] Note-se que, em particular, se Y for $\mathcal{H}$-mensurável e limitada então $\mathbb{E}\{\mathrm{XY}\,|\,\mathcal{H}\} = \mathrm{Y}\,\mathbb{E}\{\mathrm{X}\,|\,\mathcal{H}\}$ q.c.

e, em particular, quando $J = \{1, \ldots, n\}$ escreve-se,

$$\mathbb{E}(\mathrm{X} \,|\, \mathrm{Y}_1, \ldots, \mathrm{Y}_n) = \mathbb{E}\big\{\mathrm{X} \,|\, \sigma(\mathrm{Y}_1, \ldots, \mathrm{Y}_n)\big\}.$$

Uma última propriedade tem lugar quando a v.a X é independente das v.a.'s $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$:

$$\mathbb{E}(\mathrm{X} \,|\, \mathrm{Y}_1, \ldots, \mathrm{Y}_n) = \mathbb{E}(\mathrm{X}) \quad \text{q.c.}$$

Definição 1.1.20 Seja $\{\mathrm{X}_n,\, n \geqslant 1\}$ uma sucessão de v.a.'s em $\mathscr{L}_1(\Omega, \mathcal{F}, \mathbb{P})$. Dizemos que as v.a.'s X_n são **centradas nas esperanças condicionadas dados os predecessores** se,

$$\xi_n = \mathbb{E}(\mathrm{X}_n \,|\, \mathrm{X}_1, \ldots, \mathrm{X}_{n-1}) = 0 \quad \text{q.c.}$$

Apresentamos agora dois resultados fundamentais para o Capítulo 4, respectivamente, a igualdade extendida de Bienaymé e a desigualdade extendida de Kolmogorov.

Teorema 1.1.13 (igualdade extendida de Bienaymé) *Se as v.a.'s* X_n *de uma sucessão* $\{\mathrm{X}_n,\, n \geqslant 1\}$ *são centradas nas esperanças condicionadas dados os predecessores então são centradas nos valores médios e*

$$\mathbb{V}\left(\sum_{i=1}^{n} \mathrm{X}_i\right) = \sum_{i=1}^{n} \mathbb{V}(\mathrm{X}_i).$$

Demonstração. Consultar [56], página 52. ∎

Teorema 1.1.14 (desigualdade extendida de Kolmogorov) *Se* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *são centradas nas esperanças condicionadas dados os predecessores então*

$$\mathbb{P}\left\{\max_{1 \leqslant j \leqslant n} \left|\sum_{i=1}^{j} \mathrm{X}_i\right| \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon^2} \sum_{i=1}^{n} \mathbb{V}(\mathrm{X}_i).$$

Demonstração. Consultar [56], página 52. ∎

Ao longo deste texto, o termo "convergência"será utilizado como sinónimo de "existência de limite finito". Dada uma sucessão de v.a.'s $\{\mathrm{X}_n,\, n \geqslant 1\}$ em $(\Omega, \mathcal{F}, \mathbb{P})$, a sua convergência é um conceito bem definido, no sentido em que as v.a.'s X_n são funções de valores numéricos. No entanto, ao falarmos da convergência de uma sucessão de v.a.'s há uma questão técnica que sobressai de imediato: será sempre possível definir uma sucessão de v.a.'s $\{\mathrm{X}_n,\, n \geqslant 1\}$ nalgum espaço de probabilidade de modo que a função de distribuição conjunta de cada vector aleatório $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$, $n \geqslant 1$ coincida com cada função de distribuição n-dimensional $F_{1,\ldots,n}$, $n \geqslant 1$, de uma família $\{F_{1,\ldots,n},\, n \geqslant 1\}$ designada *à priori*? A resposta é afirmativa como veremos no teorema da consistência de Kolmogorov.

Definição 1.1.21 Uma família $\{F_{1,\ldots,n},\ n \geqslant 1\}$ de funções de distribuição n-dimensionais definidas para qualquer $n \geqslant 1$ é chamada **consistente** se para todo $n \geqslant 1$,

$$F_{1,\ldots,n}(x_1,\ldots,x_n) = \lim_{x_{n+1}\to\infty} F_{1,\ldots,n+1}(x_1,\ldots,x_{n+1})$$

Teorema 1.1.15 (consistência de Kolmogorov) *Se $\{F_{1,\ldots,n},\ n \geqslant 1\}$ é uma família consistente de funções de distribuição n-dimensionais então existe uma medida de probabilidade $\mathbb{P}_\infty$ sobre $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ onde $\mathbb{R}^\infty = \mathbb{R}\times\mathbb{R}\times\ldots$ tal que a função de distribuição conjunta de cada vector aleatório $(\mathrm{X}_1,\ldots,\mathrm{X}_n)$, $n \geqslant 1$ de $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \mathbb{P}_\infty)$ coincide, respectivamente, com cada função de distribuição n-dimensional $F_{1,\ldots,n}$, $n \geqslant 1$, ou seja, tal que se*

$$\mathrm{X}_m(\boldsymbol{\omega}) = \omega_m, \quad m = 1, 2, \ldots$$

para $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots) \in \mathbb{R}^\infty = \mathbb{R}\times\mathbb{R}\times\ldots$ então para todo o $n \geqslant 1$,

$$\mathbb{P}_\infty\{\mathrm{X}_1 \leqslant x_1, \ldots, \mathrm{X}_n \leqslant x_n\} = F_{1,\ldots,n}(x_1,\ldots,x_n).$$

Demonstração. Consultar [18], página 195. ∎

Definição 1.1.22 Num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ consideremos uma sucessão de v.a.'s $\{\mathrm{X}_n,\ n \geqslant 1\}$ e uma v.a. X. Dizemos que:

(C1) X_n **converge quase certamente** para X se,

$$\mathbb{P}\left\{\omega\colon\ \lim_{n\to\infty} \mathrm{X}_n(\omega) = \mathrm{X}(\omega)\right\} = 1$$

e indicamos $\mathrm{X}_n \xrightarrow{\text{q.c.}} \mathrm{X}$.

(C2) X_n **converge em probabilidade** para X se para todo o $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\{\omega\colon\ |\mathrm{X}_n(\omega) - \mathrm{X}(\omega)| > \varepsilon\} = 0$$

e escrevemos $\mathrm{X}_n \xrightarrow{\mathbb{P}} \mathrm{X}$.

Definição 1.1.23 Sendo $p > 0$, $\{\mathrm{X}_n,\ n \geqslant 1\}$ uma sucessão de v.a.'s em $\mathscr{L}_p(\Omega, \mathcal{F}, \mathbb{P})$ e X uma v.a. de $\mathscr{L}_p(\Omega, \mathcal{F}, \mathbb{P})$ dizemos que X_n **converge em momento de ordem $\boldsymbol{p}$** para X se,

$$\lim_{n\to\infty} \mathbb{E}\,|\mathrm{X}_n - \mathrm{X}|^p = 0$$

e denotamos $\mathrm{X}_n \xrightarrow{\mathscr{L}_p} \mathrm{X}$.

Estabelecemos, agora, aquele que é o tipo mais fraco de convergência de v.a.'s no sentido em que é consequência de qualquer um dos anteriores modos de convergência.

Definição 1.1.24 Sendo $\mathrm{X}, \mathrm{X}_1, \mathrm{X}_2, \ldots$ v.a.'s com, respectivamente, funções de distribuição F, F_1, $F_2, \ldots$ dizemos que X_n **converge em distribuição** ou **em lei** para X se,

$$\lim_{n\to\infty} F_n(x) = F(x)$$

para todo o ponto de continuidade x de F e escrevemos $\mathrm{X}_n \xrightarrow{\mathrm{d}} \mathrm{X}$.

Observação 1.1.6 Note-se que a convergência em distribuição é, claramente, uma propriedade exclusiva das funções distribuição das v.a.'s e não das próprias v.a.'s em si. A simbologia $\mathrm{X}_n \longrightarrow \mathrm{X}$ é simplesmente uma notação alternativa conveniente a $F_{\mathrm{X}_n} \longrightarrow F$, isto porque em geral, X é uma v.a. fictícia com f.d. F já que não está assegurado a existência de tal v.a. no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ das v.a's $\mathrm{X}_1, \mathrm{X}_2, \ldots$ (é claro que podemos sempre definir uma v.a. X com f.d. F noutro espaço de probabilidade). Inclusive, as v.a.'s $\mathrm{X}_1, \mathrm{X}_2, \ldots$ podem estar definidas em espaços de probabilidade completamente distintos (conf. Observação 1.1.1).

Uma ferramenta de grande utilidade no estudo da convergência em distribuição de sucessões de v.a.'s é a função característica,

Definição 1.1.25 Dada uma v.a. X num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ chamamos **função característica** de X à função $\varphi_{\mathrm{X}} \colon \mathbb{R} \to \mathbb{C}$ definida por,

$$\varphi_{\mathrm{X}}(t) = \mathbb{E}\left(e^{it\mathrm{X}}\right) = \mathbb{E}\big[\cos(t\mathrm{X})\big] + i\mathbb{E}\big[\sin(t\mathrm{X})\big], \quad t \in \mathbb{R}.$$

Teorema 1.1.16 *Sejam* $\mathrm{X}_1, \mathrm{X}_2, \ldots$ *v.a.'s e* $\varphi_1, \varphi_2, \ldots$, *respectivamente, as suas funções características. Se* $\mathrm{X}_n \xrightarrow{\mathrm{d}} \mathrm{X}$ *então* $\varphi_n(t) \longrightarrow \varphi(t)$ *onde* $\varphi(t)$ *é a função característica de* X. *Reciprocamente, se* $\varphi_n(t) \longrightarrow \varphi(t)$ *e a função limite* $\varphi(t)$ *é contínua em* $t = 0$ *então* $\mathrm{X}_n \xrightarrow{\mathrm{d}} \mathrm{X}$ *e* $\varphi(t)$ *é a função característica de* X.

Demonstração. Consultar [76], página 119. ∎

Mais geralmente, para um vector aleatório $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ tem-se,

Definição 1.1.26 Dado um vector aleatório $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ chamamos **função característica** de $\mathbf{X}$ à função $\varphi_{\mathbf{X}} \colon \mathbb{R}^n \to \mathbb{C}$ definida por,

$$\varphi_{\mathbf{X}}(t_1, \ldots, t_n) = \mathbb{E}\left(e^{it_1\mathrm{X}_1 + \ldots + it_n\mathrm{X}_n}\right), \quad (t_1, \ldots, t_n) \in \mathbb{R}^n.$$

O estudo da convergência quase certa de v.a.'s tem um papel absolutamente fulcral em teoria de probabilidades. Expomos, em baixo, alguns resultados úteis relativos à convergência quase certa.

Teorema 1.1.17 (lei forte de Marcinkiewicz-Zygmund) *Se* $\mathrm{X}_1, \mathrm{X}_2, \ldots$ *são variáveis aleatórias i.i.d. então para qualquer* $p \in]0, 2[$,

$$\frac{\mathrm{X}_1 + \ldots + \mathrm{X}_n - cn}{n^{1/p}} \xrightarrow{\text{q.c.}} 0$$

onde c é alguma constante finita se e só se $\mathbb{E}\big[\,|X_1|^p\,\big] < \infty$. *Mais concretamente,* $c = \mathbb{E}(X_1)$ *se* $1 \leqslant p < 2$ *e c é arbitrária se* $0 < p < 1$.

Demonstração. Consultar [18], página 125. ∎

Observação 1.1.7 Observe-se que quando $p = 1$ no Teorema 1.1.17 recuperamos a clássica **lei forte de Kolmogorov**: se $X_1, X_2, \ldots$ são variáveis aleatórias i.i.d. com $\mathbb{E}\,|X_i| < \infty$, $\forall i$ então

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{q.c.}} \mu$$

onde $\mu = \mathbb{E}(X_i)$, $\forall i$.

Teorema 1.1.18 *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias i.i.d. com* $\mathbb{E}\,|X_1| = \infty$. *Se* a_n *é uma sucessão de reais positivos tal que* a_n/n *é crescente*[11] *então,*

$$\limsup_{n\to\infty} \frac{|X_1 + \ldots + X_n|}{a_n} = 0 \quad q.c. \qquad se \qquad \sum_{n=1}^{\infty} \mathbb{P}\big\{\,|X_n| \geqslant a_n\big\} < \infty$$

ou,

$$\limsup_{n\to\infty} \frac{|X_1 + \ldots + X_n|}{a_n} = \infty \quad q.c. \qquad se \qquad \sum_{n=1}^{\infty} \mathbb{P}\big\{\,|X_n| \geqslant a_n\big\} = \infty.$$

Demonstração. Consultar [19], página 134. ∎

Teorema 1.1.19 (teorema das três séries de Kolmogorov) *Se* $\{X_n,\, n \geqslant 1\}$ *é uma sucessão de v.a.'s independentes então* $\sum_{n=1}^{\infty} X_n$ *converge q.c. se e só se*

(a) $\sum_{n=1}^{\infty} \mathbb{P}\big\{\,|X_n| > 1\big\} < \infty$.

(b) $\sum_{n=1}^{\infty} \mathbb{E}\left(X_n\, I_{\{|X_n|\leqslant 1\}}\right)$ *converge.*

(c) $\sum_{n=1}^{\infty} \mathbb{V}\left(X_n\, I_{\{|X_n|\leqslant 1\}}\right) < \infty$.

Adicionalmente, a série $\sum_{n=1}^{\infty} X_n$ *converge absolutamente q.c. se e só se* (a), (c) *e*

(b') $\sum_{n=1}^{\infty} \mathbb{E}\left|X_n\, I_{\{|X_n|\leqslant 1\}}\right| < \infty$.

estiverem asseguradas.

Demonstração. Consultar [18], páginas 117 e 122. ∎

Teorema 1.1.20 (lei forte de Feller) *Seja* $X, X_1, X_2, \ldots$ *uma sucessão de v.a.'s i.i.d. e* $\{b_n\}$ *uma sucessão positiva, crescente tal que* $\lim\limits_{n\to\infty} b_n = \infty$ *e*

$$\sum_{n=1}^{\infty} \mathbb{P}\big\{\,|X_n| > b_n\big\} < \infty.$$

Se uma das seguintes hipóteses,

[11] Uma sucessão $\{x_n\}$ diz-se crescente (resp. decrescente) se $x_n < x_{n+1}$, $\forall n \in \mathbb{N}$ (resp. $x_n > x_{n+1}$, $\forall n \in \mathbb{N}$). Do mesmo modo, se $x_n \leqslant x_{n+1}$, $\forall n \in \mathbb{N}$ (resp. $x_n \geqslant x_{n+1}$, $\forall n \in \mathbb{N}$) dizemos que a sucessão $\{x_n\}$ é não-decrescente (resp. não-crescente).

(a1) $\dfrac{b_n}{n}$ *crescente e* $\lim\limits_{n\to\infty} \dfrac{b_n}{n} = \infty$,

(a2) $\dfrac{b_n}{n}$ *decrescente,* $b_n^2 \sum\limits_{i=n}^{\infty} b_i^{-2} = O(n)$, $n \to \infty$ *e* $\mathbb{E}(\mathrm{X}) = 0$,

(a3) $b_n^2 \sum\limits_{i=n}^{\infty} b_i^{-2} = O(n)$, $n \to \infty$, $b_n \sum\limits_{i=1}^{n} b_i^{-1} = O(n)$, $n \to \infty$ *e* $\mathbb{E}(\mathrm{X}) = 0$,

é verificada então $\dfrac{1}{b_n} \sum\limits_{i=1}^{n} \mathrm{X}_i \xrightarrow{\text{q.c.}} 0$.

Demonstração. Consultar [18], página 126. ∎

Teorema 1.1.21 *Seja* $\{\mathrm{X}_n,\, n \geqslant 1\}$ *uma sucessão de v.a.'s tal que,*

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{\infty} \mathbb{P}\big\{ |\mathrm{X}_n| > \varepsilon \big\} < \infty \tag{1.1.5}$$

então $\mathrm{X}_n \xrightarrow{\text{q.c.}} 0$. *Se as v.a.'s* $\mathrm{X}_1, \mathrm{X}_2, \ldots$ *forem independentes e* $\mathrm{X}_n \xrightarrow{\text{q.c.}} 0$ *então a condição* (1.1.5) *fica assegurada.*

Demonstração. Consultar [74], página 271. A demonstração da segunda asserção pode ser vista em [55], página 240. ∎

1.2 Elementos de análise matricial

Nesta secção faremos uma breve exposição de alguns conceitos elementares de análise matricial com o objectivo de sustentar resultados vindouros. Por esta razão, daremos especial ênfase às matrizes rectangulares com entradas reais, mas a generalização às matrizes rectangulares sobre qualquer corpo (de escalares) $\mathbb{K}$ pode ser sempre efectuada (ver por exemplo [40]). No que diz respeito aos espaços vectoriais, elegemos o espaço vectorial $\mathbb{R}^n$ sobre o corpo de escalares $\mathbb{R}$ equipado com a base canónica para realizar o nosso trabalho, mas a extensão a qualquer espaço vectorial $\mathbb{E}$ sobre um corpo $\mathbb{K}$ munido de uma outra base é naturalmente possível. Neste sentido recomendamos o excelente [68] ou ainda [59]. A fechar esta secção apresentaremos o importante *teorema de Cochran* cuja a demonstração pode ser encontrada em [21].

Dá-se o nome de **matriz rectangular** do tipo $n \times \kappa$ sobre $\mathbb{R}$ a um quadro onde $n\kappa$ elementos de $\mathbb{R}$ se dispõem de modo a formarem n filas horizontais de κ elementos cada uma (as **linhas da matriz**) e κ filas verticais de n elementos cada uma (as **colunas da matriz**),

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1\kappa} \\ a_{21} & a_{22} & \ldots & a_{2\kappa} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{n\kappa} \end{bmatrix}.$$

Abreviadamente, escrevemos $[a_{ij}]_{\substack{i=1,2,\dots,n\\ j=1,2,\dots,\kappa}}$ e representamos por $\mathcal{M}_{n\times\kappa}(\mathbb{R})$ o conjunto de todas as matrizes do tipo $n\times\kappa$ sobre o corpo $\mathbb{R}$. Uma matriz do tipo $n\times n$ diz-se **quadrada** de ordem n; chamamos ainda **vector coluna** a uma matriz do tipo $n\times 1$ e **vector linha** a uma matriz do tipo $1\times\kappa$.

EXEMPLO 1.2.1

A matriz $[0]_{\substack{i=1,\dots,n\\ j=1,\dots,n}}\in\mathcal{M}_{n\times n}(\mathbb{R})$ cujas entradas são todas nulas chama-se **matriz nula** e será denotada por $\mathbf{O}$. Uma matriz

$$\boldsymbol{D}=\begin{bmatrix} d_{11} & 0 & \dots & 0\\ 0 & d_{22} & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & d_{nn}\end{bmatrix}.$$

de $\mathcal{M}_{n\times n}(\mathbb{R})$ chama-se **matriz diagonal** e representa-se por $\boldsymbol{D}=\operatorname{diag}(d_{11},\dots,d_{nn})$; um caso particular de uma matriz diagonal é a **matriz identidade** onde $d_{ii}=1$ para todo o $i=1,\dots,n$ e que é habitualmente indicada por $\boldsymbol{I}$.

Se $\boldsymbol{A}=[a_{ij}]_{\substack{i=1,\dots,n\\ j=1,\dots,\kappa}}\in\mathcal{M}_{n\times\kappa}(\mathbb{R})$ chamamos **matriz transposta** de $\boldsymbol{A}$ à matriz $[a_{ji}]_{\substack{j=1,\dots,\kappa\\ i=1,\dots,n}}\in\mathcal{M}_{\kappa\times n}(\mathbb{R})$ habitualmente denotada por $\boldsymbol{A}^T$. Uma matriz $\boldsymbol{A}\in\mathcal{M}_{n\times n}$ diz-se **simétrica** se $\boldsymbol{A}^T=\boldsymbol{A}$.

Um importante inteiro não-negativo associado a cada matriz $\boldsymbol{A}\in\mathcal{M}_{n\times\kappa}(\mathbb{R})$ é a **característica** de $\boldsymbol{A}$ denotada usualmente por $\operatorname{car}(\boldsymbol{A})$. Se $\boldsymbol{A}\in\mathcal{M}_{n\times\kappa}(\mathbb{R})$, $\operatorname{car}(\boldsymbol{A})$ é o número máximo de vectores coluna linearmente independentes de $\boldsymbol{A}$. Um facto notável é a veracidade da igualdade $\operatorname{car}(\boldsymbol{A}^T)=\operatorname{car}(\boldsymbol{A})$; de um modo geral, tem-se $0\leqslant\operatorname{car}(\boldsymbol{A})\leqslant\min(n,\kappa)$.

A cada aplicação linear $T\colon\mathbb{R}^\kappa\longrightarrow\mathbb{R}^n$ está associada uma matriz do tipo $n\times\kappa$ definida do seguinte modo: supondo os vectores da base canónica de $\mathbb{R}^\kappa$ tem-se,

$$\begin{aligned} T(1,0,\dots,0)&=(t_{11},t_{21},\dots,t_{n1})\\ T(0,1,\dots,0)&=(t_{12},t_{22},\dots,t_{n2})\\ &\vdots\\ T(0,0,\dots,1)&=(t_{1\kappa},t_{2\kappa},\dots,t_{n\kappa}).\end{aligned}$$

e a matriz da aplicação linear T fica definida escrevendo $\boldsymbol{T}=[t_{ij}]_{\substack{i=1,2,\dots,n\\ j=1,2,\dots,\kappa}}$. Reciprocamente, para cada matriz $\boldsymbol{A}$ do tipo $n\times\kappa$ existe exactamente uma aplicação linear $T\colon\mathbb{R}^\kappa\longrightarrow\mathbb{R}^n$ tal que $\boldsymbol{T}=\boldsymbol{A}$ (os vectores $T(1,0,\dots,0),T(0,1,\dots,0),\dots,T(0,0,\dots,1)$ determinam univocamente a aplicação T). As operações de adição, multiplicação e multiplicação por um escalar real de matrizes são definidas de forma a serem compatíveis, respectivamente, com a adição, multiplicação (composição) e multiplicação por um escalar real de aplicações lineares. Por outras palavras, estas operações estão definidas como se segue:

(A) Se $[a_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ e $[b_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ então,

$$[a_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}} + [b_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}} = [a_{ij}+b_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}}$$

(M) Se $[a_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ e $[c_{ij}]_{\substack{i=1,\ldots,\kappa\\ j=1,\ldots,m}} \in \mathcal{M}_{\kappa\times m}(\mathbb{R})$ então,

$$[a_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}} [c_{ij}]_{\substack{i=1,\ldots,\kappa\\ j=1,\ldots,m}} = \left[\sum_{p=1}^{n} a_{ip}c_{pj}\right]_{\substack{i=1,\ldots,n\\ j=1,\ldots,m}}$$

(m) Se $[a_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ e $\lambda \in \mathbb{R}$ então,

$$\lambda\,[a_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}} = [\lambda a_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}}$$

Uma matriz quadrada $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ é **invertível** quando existe uma matriz quadrada $\boldsymbol{X}$ de ordem n tal que $\boldsymbol{AX} = \boldsymbol{I} = \boldsymbol{XA}$ onde $\boldsymbol{I}$ é a matriz identidade. Quando uma matriz quadrada $\boldsymbol{A}$ de ordem n é invertível, chamamos **matriz inversa** à matriz quadrada $\boldsymbol{X}$ de ordem n que satisfaz $\boldsymbol{AX} = \boldsymbol{I} = \boldsymbol{XA}$; a matriz $\boldsymbol{X}$ representa-se habitualmente por $\boldsymbol{A}^{-1}$. Se $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ for uma matriz invertível então

$$\left(\boldsymbol{A}^T\right)^{-1} = \left(\boldsymbol{A}^{-1}\right)^T. \tag{1.2.1}$$

Se $\boldsymbol{A} = [a_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,n}}$ for uma matriz arbitrária de $\mathcal{M}_{n\times n}(\mathbb{R})$ podemos considerar a função que a cada vector coluna $\boldsymbol{x} = [x_1\ \ldots\ x_n]^T$ faz corresponder o valor real,

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \sum_{i,j=1}^{n} a_{ij}x_ix_j.$$

Uma função de $\boldsymbol{x}$ que seja expressa na forma $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ designa-se por **forma quadrática** em $\boldsymbol{x}$ e é usual referir-mo-nos à matriz $\boldsymbol{A}$ como a **matriz da forma quadrática $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$**.

Dizemos que uma matriz $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ é **definida positiva** se $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ para todo o vector coluna $\boldsymbol{x} \in \mathcal{M}_{n\times 1}(\mathbb{R})$ não-nulo. Dizemos ainda que $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ é **definida não-negativa** se $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geqslant 0$ para todo o vector coluna $\boldsymbol{x} \in \mathcal{M}_{n\times 1}(\mathbb{R})$. Uma matriz $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ que seja definida não-negativa mas não definida positiva (i.e. $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geqslant 0$ para todo o vector coluna $\boldsymbol{x} \in \mathcal{M}_{n\times 1}(\mathbb{R})$ e $\boldsymbol{x}_0^T\boldsymbol{A}\boldsymbol{x}_0 = 0$ para algum vector coluna $\boldsymbol{x}_0 \in \mathcal{M}_{n\times 1}(\mathbb{R})$ não-nulo) chama-se **semidefinida positiva**.[12]

Teorema 1.2.1 *Sejam* $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m \in \mathcal{M}_{n\times n}(\mathbb{R})$.

(a) *Se* $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ *são definidas não-negativas então* $\boldsymbol{A}_1 + \ldots + \boldsymbol{A}_m$ *é definida não-negativa.*

[12] De referir que, uma forma quadrática $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ diz-se **definida não-negativa** quando $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geqslant 0$ para todo o vector coluna $\boldsymbol{x}$; uma forma quadrática $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ diz-se ainda **definida positiva** quando $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ para todo o vector coluna $\boldsymbol{x}$ não-nulo. Uma forma quadrática que seja definida não-negativa, mas que não seja definida positiva, diz-se **semidefinida positiva**.

(b) *Se pelo menos uma das matrizes* $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ *é definida positiva e as restantes definidas não-negativas então* $\boldsymbol{A}_1 + \ldots + \boldsymbol{A}_m$ *é definida positiva.*

Demonstração. Consultar [38], página 212. ∎

Teorema 1.2.2 *Se* $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ *é uma matriz definida positiva então* $\boldsymbol{A}^{-1}$ *também é definida positiva. Se* $\boldsymbol{A}$ *for semidefinida positiva e invertível então a sua inversa* $\boldsymbol{A}^{-1}$ *é semidefinida positiva.*

Demonstração. Consultar [38], página 214. ∎

Teorema 1.2.3 *Se* $\boldsymbol{A} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ *então a matriz* $\boldsymbol{A}^T\boldsymbol{A}$ *é definida não-negativa. Além disso, se* $\operatorname{car}(\boldsymbol{A}) = \kappa$ *então* $\boldsymbol{A}^T\boldsymbol{A}$ *é definida positiva; caso contrário, se* $\operatorname{car}(\boldsymbol{A}) < \kappa$ *então* $\boldsymbol{A}^T\boldsymbol{A}$ *é semidefinida positiva.*

Demonstração. Consultar [38], página 214. ∎

Daqui para a frente, por convenção, consideraremos os elementos de $\mathbb{R}^n$ como vectores coluna, ou seja, $\mathbb{R}^n = \mathcal{M}_{n\times 1}(\mathbb{R})$. Pelo que já foi descrito anteriormente, também olharemos sempre para uma matriz $\boldsymbol{A} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ como uma transformação linear de $\mathbb{R}^\kappa$ para $\mathbb{R}^n$ (em relação às bases canónicas de $\mathbb{R}^\kappa$ e $\mathbb{R}^n$); o seu **espaço imagem** será o subespaço vectorial $\operatorname{Im}(\boldsymbol{A}) = \{\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^n\colon\ \boldsymbol{x} \in \mathbb{R}^\kappa\}$ de $\mathbb{R}^n$ e o **núcleo** de $\boldsymbol{A}$ estará definido como sendo o subespaço vectorial de $\mathbb{R}^\kappa$ dado por $\operatorname{Nuc}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^\kappa\colon \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$.

Observação 1.2.1 O espaço imagem de uma matriz $\boldsymbol{A} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ é o subespaço vectorial de $\mathbb{R}^n$ gerado pelas colunas de $\boldsymbol{A}$.

Teorema 1.2.4 *Se* $\boldsymbol{A} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ *então* $\operatorname{car}(\boldsymbol{A}) = \dim\big(\operatorname{Im}(\boldsymbol{A})\big)$.

Demonstração. Consultar [68], página 163. ∎

O núcleo e o espaço imagem estão relacionados do seguinte modo,

$$\operatorname{Nuc}(\boldsymbol{A}^T) = \big(\operatorname{Im}(\boldsymbol{A})\big)^{\perp} \quad \text{e} \quad \operatorname{Im}(\boldsymbol{A}^T) = \big(\operatorname{Nuc}(\boldsymbol{A})\big)^{\perp} \tag{1.2.2}$$

onde definimos o **complemento ortogonal** de um qualquer subconjunto $\mathbb{F} \in \mathbb{R}^n$ como sendo o subespaço vectorial,

$$\mathbb{F}^{\perp} = \{\boldsymbol{x} \in \mathbb{R}^n\colon\ \boldsymbol{x}^T\boldsymbol{y} = 0,\ \text{para todo o } \boldsymbol{y} \in \mathbb{F}\}.$$

O escalar $\boldsymbol{x}^T\boldsymbol{y}$ é vulgarmente chamado de **produto interno canónico** de $\boldsymbol{x} \in \mathbb{R}^n$ e $\boldsymbol{y} \in \mathbb{R}^n$ e é muitas vezes denotado por $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$. Recordemos que, mais geralmente, um **produto interno** definido num espaço vectorial $\mathbb{E}$ sobre $\mathbb{R}$ é uma aplicação $h\colon \mathbb{E}\times\mathbb{E} \longrightarrow \mathbb{R}$ que satisfaz, para quaisquer $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{E}$ e para todo o $\lambda \in \mathbb{R}$,

(p1) $h(\boldsymbol{x}, \boldsymbol{y}) = h(\boldsymbol{y}, \boldsymbol{x})$.

(p2) $h(\boldsymbol{x}+\boldsymbol{y},\boldsymbol{z}) = h(\boldsymbol{x},\boldsymbol{z}) + h(\boldsymbol{y},\boldsymbol{z})$.

(p3) $h(\lambda\boldsymbol{x},\boldsymbol{y}) = \lambda h(\boldsymbol{x},\boldsymbol{y})$.

(p4) Se $\boldsymbol{x} \neq \boldsymbol{0}$ então $h(\boldsymbol{x},\boldsymbol{x}) > 0$.

Naturalmente, a aplicação $\langle\, \cdot\,,\, \cdot\,\rangle\colon \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ dada por,

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle = x_1y_1 + \ldots + x_ny_n, \quad \boldsymbol{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n, \quad \boldsymbol{y} = (y_1,\ldots,y_n) \in \mathbb{R}^n$$

e definida anteriormente como produto interno canónico verifica os axiomas (p1), (p2), (p3) e (p4). Convém referir que, ao longo deste texto, assumiremos sempre o espaço vectorial $\mathbb{R}^n$ munido do produto interno canónico. Os vectores $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$ dizem-se **ortogonais** se $\langle \boldsymbol{x},\boldsymbol{y}\rangle = 0$ e indicamos $\boldsymbol{x} \perp \boldsymbol{y}$. Dado um vector $\boldsymbol{x} \in \mathbb{R}^n$ chamamos **norma euclideana** de $\boldsymbol{x} \in \mathbb{R}^n$ ao escalar não-negativo $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x},\boldsymbol{x}\rangle} = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}$. Uma matriz $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ diz-se **ortogonal** se $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I} = \boldsymbol{A}\boldsymbol{A}^T$.[13] Dada uma matriz $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ ortogonal então $\boldsymbol{A}$ é invertível e $\boldsymbol{A}^{-1} = \boldsymbol{A}^T$. Se $\mathbb{F}$ for um subespaço vectorial de $\mathbb{R}^n$ então $\mathbb{R}^n = \mathbb{F} \oplus \mathbb{F}^\perp$, ou seja, $\mathbb{R}^n$ é soma directa de $\mathbb{F}$ e $\mathbb{F}^\perp$ e, portanto, cada elemento $\boldsymbol{x} \in \mathbb{R}^n$ pode ser representado univocamente como a soma de um elemento de $\mathbb{F}$ com um elemento de $\mathbb{F}^\perp$,

$$\boldsymbol{x} = \boldsymbol{x}_{\mathbb{F}} + \boldsymbol{x}_{\mathbb{F}^\perp} \text{ onde } \boldsymbol{x}_{\mathbb{F}} \in \mathbb{F} \text{ e } \boldsymbol{x}_{\mathbb{F}^\perp} \in \mathbb{F}^\perp.$$

À aplicação linear $P_{\mathbb{F}} : \mathbb{R}^n \longrightarrow \mathbb{F}$ dada por $P_{\mathbb{F}}(\boldsymbol{x}) = \boldsymbol{x}_{\mathbb{F}}$ chamamos **projecção ortogonal** de $\mathbb{R}^n$ sobre $\mathbb{F}$; à aplicação linear $P_{\mathbb{F}^\perp} = Id - P_{\mathbb{F}}$ dada por $P_{\mathbb{F}^\perp}(\boldsymbol{x}) = (Id - P_{\mathbb{F}})(\boldsymbol{x}) = \boldsymbol{x} - P_{\mathbb{F}}(\boldsymbol{x}) = \boldsymbol{x}_{\mathbb{F}^\perp}$ chamamos projecção ortogonal de $\mathbb{R}^n$ sobre $\mathbb{F}^\perp$. Se $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_\kappa\}$ for uma base ortonormada[14] de $\mathbb{F}$ então,

$$\boldsymbol{P}_{\mathbb{F}}\boldsymbol{x} = \sum_{i=1}^{\kappa} \langle \boldsymbol{x},\boldsymbol{v}_i\rangle \boldsymbol{v}_i \text{ para todo o } \boldsymbol{x} \in \mathbb{R}^n \quad \text{e} \quad \|\boldsymbol{P}_{\mathbb{F}}\boldsymbol{x}\|^2 = \sum_{i=1}^{\kappa} \langle \boldsymbol{x},\boldsymbol{v}_i\rangle^2 \tag{1.2.3}$$

onde $\boldsymbol{P}_{\mathbb{F}}$ é a matriz associada à aplicação linear $P_{\mathbb{F}}$. Enfim, dado $\boldsymbol{x} \in \mathbb{R}^n$ é ainda válida a conhecida **fórmula de Pitágoras**,

$$\|\boldsymbol{x}\|^2 = \|\boldsymbol{P}_{\mathbb{F}}\boldsymbol{x}\|^2 + \|\boldsymbol{P}_{\mathbb{F}^\perp}\boldsymbol{x}\|^2$$

em que $\boldsymbol{P}_{\mathbb{F}^\perp}$ designa a matriz associada à aplicação linear $P_{\mathbb{F}^\perp}$.

Um método clássico de ortogonalização de um conjunto de vectores linearmente independentes $\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_p\}$ de $\mathbb{R}^n$ é o **processo de ortogonalização de Gram-Schmidt**. O algoritmo que

[13] Voltando ainda à associação entre as aplicações lineares e as matrizes, convém sublinhar que o termo **transformação ortogonal** utilizado mais adiante (ver teorema de Cochran) refere-se a qualquer aplicação linear $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ cuja matriz associada $\boldsymbol{Q} \in \mathcal{M}_{n\times n}(\mathbb{R})$ é ortogonal.

[14] Recordemos que um conjunto de vectores $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_\kappa\} \subset \mathbb{R}^n$ diz-se ortonormado se $\|\boldsymbol{v}_1\| = \ldots = \|\boldsymbol{v}_\kappa\| = 1$ e $\boldsymbol{v}_i \perp \boldsymbol{v}_j$ para todo $i \neq j$. Uma base ortonormada de um subespaço vectorial $\mathbb{F}$ de $\mathbb{R}^n$ com dimensão κ é um conjunto de κ vectores de $\mathbb{F}$ ortonormado (observe-se que qualquer conjunto de vectores ortonormado é automaticamente linearmente independente).

passaremos a descrever, permite encontrar um conjunto ortogonal de vectores[15] (não-nulos) à custa do conjunto de vectores linearmente independentes $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\}$ de $\mathbb{R}^n$ da seguinte maneira: definindo recursivamente,

$$
\begin{aligned}
\boldsymbol{y}_1 &= \boldsymbol{x}_1 \\
\boldsymbol{y}_2 &= \boldsymbol{x}_2 - \frac{\langle \boldsymbol{x}_2, \boldsymbol{y}_1 \rangle}{\langle \boldsymbol{y}_1, \boldsymbol{y}_1 \rangle} \boldsymbol{y}_1 \\
\boldsymbol{y}_3 &= \boldsymbol{x}_3 - \frac{\langle \boldsymbol{x}_3, \boldsymbol{y}_1 \rangle}{\langle \boldsymbol{y}_1, \boldsymbol{y}_1 \rangle} \boldsymbol{y}_1 - \frac{\langle \boldsymbol{x}_3, \boldsymbol{y}_2 \rangle}{\langle \boldsymbol{y}_2, \boldsymbol{y}_2 \rangle} \boldsymbol{y}_2 \\
&\;\;\vdots \\
\boldsymbol{y}_p &= \boldsymbol{x}_p - \frac{\langle \boldsymbol{x}_p, \boldsymbol{y}_1 \rangle}{\langle \boldsymbol{y}_1, \boldsymbol{y}_1 \rangle} \boldsymbol{y}_1 - \frac{\langle \boldsymbol{x}_p, \boldsymbol{y}_2 \rangle}{\langle \boldsymbol{y}_2, \boldsymbol{y}_2 \rangle} \boldsymbol{y}_2 - \ldots - \frac{\langle \boldsymbol{x}_p, \boldsymbol{y}_{p-1} \rangle}{\langle \boldsymbol{y}_{p-1}, \boldsymbol{y}_{p-1} \rangle} \boldsymbol{y}_{p-1}
\end{aligned}
$$

obtém-se o conjunto ortogonal de vectores $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p\}$ de $\mathbb{R}^n$. Convém destacar que este processo conduz-nos automaticamente a um conjunto ortonormado de vectores de $\mathbb{R}^n$, a saber, $\left\{\frac{\boldsymbol{y}_1}{\|\boldsymbol{y}_1\|}, \ldots, \frac{\boldsymbol{y}_p}{\|\boldsymbol{y}_p\|}\right\}$ revelando-se, assim, um método de grande utilidade na construção de bases ortonormadas em espaços vectoriais. De referir ainda, que o processo de ortogonalização de Gram-Schmidt pode também ser utilizado em conjuntos finitos ou infinitos numeráveis de vectores não necessariamente linearmente independentes de um espaço vectorial[16] (se o conjunto não for linearmente independente o método produzirá um vector $\boldsymbol{y}_m = \boldsymbol{0}$ onde m é o menor valor que torna o conjunto $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}$ linearmente dependente).

Proposição 1.2.1 *Se $\boldsymbol{A} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ uma matriz com característica κ então existe uma base ortonormada $\{\boldsymbol{w}_1(n), \ldots, \boldsymbol{w}_n(n)\}$ de $\mathbb{R}^n$ tal que,*

$$\left\|\boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A})}\boldsymbol{x}\right\|^2 = \langle \boldsymbol{w}_1(n), \boldsymbol{x} \rangle^2 + \ldots + \langle \boldsymbol{w}_\kappa(n), \boldsymbol{x} \rangle^2, \quad \boldsymbol{x} \in \mathbb{R}^n.$$

Demonstração. Visto que a matriz $\boldsymbol{A}_n = [a_{ij}]_{\substack{i=1,\ldots,n\\ j=1,\ldots,\kappa}}$ tem característica κ podemos considerar uma base de $\mathrm{Im}(\boldsymbol{A})$ dada por,

$$
\begin{aligned}
\boldsymbol{a}_1 &= (a_{11}, a_{21}, \ldots, a_{n1}) \\
&\;\;\vdots \\
\boldsymbol{a}_\kappa &= (a_{1\kappa}, a_{2\kappa}, \ldots, a_{n\kappa}).
\end{aligned}
$$

Utilizando o processo de ortogonalização de Gram-Schmidt podemos construir (primeiro uma base ortogonal e depois) uma base ortonormada $\{\boldsymbol{w}_1(n), \ldots, \boldsymbol{w}_\kappa(n)\}$ de $\mathrm{Im}(\boldsymbol{A})$ tendo-se,

$$\left\|\boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A})}\boldsymbol{x}\right\|^2 = \langle \boldsymbol{w}_1(n), \boldsymbol{x} \rangle^2 + \ldots + \langle \boldsymbol{w}_\kappa(n), \boldsymbol{x} \rangle^2, \quad \boldsymbol{x} \in \mathbb{R}^n$$

(conf. (1.2.3)).

[15]Um conjunto de vectores $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p\} \subset \mathbb{R}^n$ diz-se ortogonal se cada par de vectores distintos de $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p\}$ é ortogonal, ou seja, se $\boldsymbol{y}_i \perp \boldsymbol{y}_j$ para todo o $i \neq j$.

[16]Um conjunto de vectores ortogonal infinito numerável será produzido, num espaço de dimensão infinita, a partir de um conjunto infinito numerável de vectores linearmente independentes.

Do mesmo modo, o complemento ortogonal $\big(\mathrm{Im}(\boldsymbol{A})\big)^{\perp}$ de $\mathrm{Im}(\boldsymbol{A})$ admitirá também uma base ortonormada $\big\{\boldsymbol{w}_{\kappa+1}(n),\ldots,\boldsymbol{w}_n(n)\big\}$ e os vectores

$$\begin{aligned}
\boldsymbol{w}_1(n) &= \big(w_{11}(n), w_{21}(n), \ldots, w_{n1}(n)\big)\\
&\ \ \vdots\\
\boldsymbol{w}_\kappa(n) &= \big(w_{1\kappa}(n), w_{2\kappa}(n), \ldots, w_{n\kappa}(n)\big)\\
\boldsymbol{w}_{\kappa+1}(n) &= \big(w_{1\kappa+1}(n), w_{2\kappa+1}(n), \ldots, w_{n\kappa+1}(n)\big)\\
&\ \ \vdots\\
\boldsymbol{w}_n(n) &= \big(w_{1n}(n), w_{2n}(n), \ldots, w_{nn}(n)\big)
\end{aligned}$$

constituirão uma base ortonormada de $\mathbb{R}^n = \mathrm{Im}(\boldsymbol{A}) \oplus \big(\mathrm{Im}(\boldsymbol{A})\big)^{\perp}$. ∎

Um vector $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v} \neq \boldsymbol{0}$ diz-se **vector próprio** de $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ se existir um escalar $\lambda \in \mathbb{R}$ tal que $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$. O escalar λ toma o nome de **valor próprio** de $\boldsymbol{A}$ associado ao vector próprio $\boldsymbol{v}$. Ao conjunto,

$$\mathrm{Spec}(\boldsymbol{A}) = \big\{\lambda \in \mathbb{R} :\ \lambda \text{ é valor próprio de } \boldsymbol{A}\big\}$$

de todos os valores próprios de $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ damos o nome de **espectro** de $\boldsymbol{A}$. O **raio espectral** de $\boldsymbol{A}$ é o número real não-negativo $\rho(\boldsymbol{A}) = \max\big\{\,|\lambda| :\ \lambda \in \mathrm{Spec}(\boldsymbol{A})\big\}$.

Teorema 1.2.5 *Se λ é um valor próprio de $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ associado ao vector próprio $\boldsymbol{v}$ então, para qualquer inteiro positivo m, λ^m é valor próprio $\boldsymbol{A}^m$ associado ao vector próprio $\boldsymbol{v}$. Além disso, se $\boldsymbol{A}$ for invertível (e nesse caso $\lambda \neq 0$) então λ^{-1} é valor próprio de $\boldsymbol{A}^{-1}$ associado ao vector próprio $\boldsymbol{v}$.*

Demonstração. Consultar [38], página 521. ∎

Teorema 1.2.6 *Seja $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ uma matriz simétrica. Então,*

(a) *$\boldsymbol{A}$ é definida não-negativa se e só se todos os seus valores próprios são não-negativos.*

(b) *$\boldsymbol{A}$ é definida positiva se e só se todos os seus valores próprios são positivos.*

(c) *$\boldsymbol{A}$ é semidefinida positiva se e só se todos os seus valores próprios são não-negativos e pelo menos um deles é nulo.*

Demonstração. Consultar [38], página 543. ∎

É do conhecimento geral que toda a matriz $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ simétrica admite uma diagonalização ortogonal, ou seja, existe uma matriz ortogonal $\boldsymbol{Q} \in \mathcal{M}_{n\times n}(\mathbb{R})$ tal que $\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q} = \mathrm{diag}(\lambda_1,\ldots,\lambda_n)$ onde $\lambda_1,\ldots,\lambda_n$ representam os não necessariamente distintos valores próprios de $\boldsymbol{A}$ ordenados de forma arbitrária (ver [38]). Assim, vamos assumir que os elementos $\lambda_1,\ldots,\lambda_n$ estão ordenados de forma não-decrescente,

$$\lambda_{\min} = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_{n-1} \leqslant \lambda_n = \lambda_{\max}.$$

O menor e o maior valor próprio de uma matriz simétrica $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ são caracterizados pela solução de problemas de mínimo e máximo com constrangimentos.

Teorema 1.2.7 (teorema de Rayleigh-Ritz) *Se* $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ *é uma matriz simétrica com valores próprios* $\lambda_1, \ldots, \lambda_n$ *(não necessariamente distintos) ordenados de forma não-decrescente* $\lambda_{\min} = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_{n-1} \leqslant \lambda_n = \lambda_{\max}$ *então,*

$$\lambda_{\min}\boldsymbol{x}^T\boldsymbol{x} \leqslant \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \leqslant \lambda_{\max}\boldsymbol{x}^T\boldsymbol{x}, \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Adicionalmente,

$$\lambda_{\max} = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} \quad e \quad \lambda_{\min} = \min_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}.$$

Demonstração. Consultar [40], página 176. ∎

Proposição 1.2.2 *Se* $\boldsymbol{A} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ *então* $\rho(\boldsymbol{A}\boldsymbol{A}^T) = \rho(\boldsymbol{A}^T\boldsymbol{A})$.

Demonstração. As matrizes $\boldsymbol{A}\boldsymbol{A}^T$ e $\boldsymbol{A}^T\boldsymbol{A}$ são simétricas e definidas não-negativas donde, pelo Teorema 1.2.6, os valores próprios de ambas são números reais não-negativos.

1. $\rho(\boldsymbol{A}\boldsymbol{A}^T) = 0$.

Com vista a um absurdo, suponhamos que $\rho(\boldsymbol{A}^T\boldsymbol{A}) > 0$. Então existe um vector $\boldsymbol{v} \neq \boldsymbol{0}$ de $\mathbb{R}^\kappa$ tal que $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v} = \rho(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{v}$ e necessariamente $\boldsymbol{A}\boldsymbol{v} \neq \boldsymbol{0}$ (pois estamos a assumir que $\rho(\boldsymbol{A}^T\boldsymbol{A}) > 0$). Como,

$$\boldsymbol{A}\boldsymbol{v} \neq \boldsymbol{0} \quad \text{e} \quad \boldsymbol{A}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{v}) = \rho(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{A}\boldsymbol{v}$$

tem-se que $\rho(\boldsymbol{A}^T\boldsymbol{A})$ é valor próprio $\boldsymbol{A}\boldsymbol{A}^T$ donde $0 < \rho(\boldsymbol{A}^T\boldsymbol{A}) \leqslant \rho(\boldsymbol{A}\boldsymbol{A}^T) = 0$ o que é absurdo. Portanto $\rho(\boldsymbol{A}^T\boldsymbol{A}) = 0$.

2. $\rho(\boldsymbol{A}\boldsymbol{A}^T) > 0$.

Se $\rho(\boldsymbol{A}\boldsymbol{A}^T) > 0$ existirá um vector $\boldsymbol{w} \neq \boldsymbol{0}$ de $\mathbb{R}^n$ tal que $\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{w} = \rho(\boldsymbol{A}\boldsymbol{A}^T)\boldsymbol{w}$ e necessariamente $\boldsymbol{A}^T\boldsymbol{w} \neq \boldsymbol{0}$ (porque $\rho(\boldsymbol{A}\boldsymbol{A}^T) > 0$). Visto que,

$$\boldsymbol{A}^T\boldsymbol{w} \neq \boldsymbol{0} \quad \text{e} \quad \boldsymbol{A}^T\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{w}) = \rho(\boldsymbol{A}\boldsymbol{A}^T)\boldsymbol{A}^T\boldsymbol{w}$$

então $\rho(\boldsymbol{A}\boldsymbol{A}^T)$ é valor próprio $\boldsymbol{A}^T\boldsymbol{A}$ donde

$$0 < \rho(\boldsymbol{A}\boldsymbol{A}^T) \leqslant \rho(\boldsymbol{A}^T\boldsymbol{A}). \tag{1.2.4}$$

Por outro lado, também $\rho(\boldsymbol{A}^T\boldsymbol{A}) > 0$ pois se $\rho(\boldsymbol{A}^T\boldsymbol{A}) = 0$ então um raciocínio (por absurdo) análogo ao utilizado em **1.** conduziria a $\rho(\boldsymbol{A}\boldsymbol{A}^T) = 0$. Mas neste caso, já sabemos (de **1.**) que,

$$0 < \rho(\boldsymbol{A}^T\boldsymbol{A}) \leqslant \rho(\boldsymbol{A}\boldsymbol{A}^T)$$

o que em combinação com (1.2.4) produz $\rho(\boldsymbol{A}\boldsymbol{A}^T) = \rho(\boldsymbol{A}^T\boldsymbol{A})$. ∎

Observação 1.2.2 Observemos que se $\boldsymbol{A} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ é não-nula então as matrizes $\boldsymbol{A}\boldsymbol{A}^T$ e $\boldsymbol{A}^T\boldsymbol{A}$ são semidefinidas positivas e têm os mesmos valores próprios positivos o que implica $\rho(\boldsymbol{A}\boldsymbol{A}^T) = \rho(\boldsymbol{A}^T\boldsymbol{A})$.

Os dois próximos resultados são estimativas matriciais fundamentais nos capítulos que se vão seguir.

Proposição 1.2.3 *Se* $\boldsymbol{A} \in \mathcal{M}_{n\times\kappa}(\mathbb{R})$ *então*

$$\left\| \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{x} \right\|^2 \leqslant \rho\Big(\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\Big) \left\| \boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A})}\boldsymbol{x} \right\|^2, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$$

onde $\boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A})}$ *é a matriz associada à projecção ortogonal sobre o espaço imagem* $\mathrm{Im}(\boldsymbol{A})$.

Demonstração. Escrevendo $\mathbb{R}^n = \mathrm{Im}(\boldsymbol{A}) \oplus \left(\mathrm{Im}(\boldsymbol{A})\right)^{\perp}$ tem-se,

$$\mathbb{R}^n = \mathrm{Im}(\boldsymbol{A}) \oplus \mathrm{Nuc}(\boldsymbol{A}^T)$$

(conf. (1.2.2)) e qualquer vector $\boldsymbol{x} \in \mathbb{R}^n$ pode ser decomposto de forma única em,

$$\boldsymbol{x} = \boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} + \boldsymbol{x}_{\mathrm{Nuc}(\boldsymbol{A}^T)}$$

onde $\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} \in \mathrm{Im}(\boldsymbol{A})$ e $\boldsymbol{x}_{\mathrm{Nuc}(\boldsymbol{A}^T)} \in \mathrm{Nuc}(\boldsymbol{A}^T)$. Então $\boldsymbol{A}^T\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} + \boldsymbol{A}^T\boldsymbol{x}_{\mathrm{Nuc}(\boldsymbol{A}^T)} = \boldsymbol{A}^T\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})}$ e uma vez que $\left[\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\right]\left[\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\right]$ é simétrica definida não-negativa, todos os seus valores próprios são não-negativos tendo-se,

$$\begin{aligned}
\left\| \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{x} \right\|^2 &= \left\| \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} \right\|^2 \\
&= \boldsymbol{x}^T_{\mathrm{Im}(\boldsymbol{A})}\left[\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\right]^T\left[\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\right]\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} \\
&= \boldsymbol{x}^T_{\mathrm{Im}(\boldsymbol{A})}\left[\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\right]\left[\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\right]\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} \\
&\underset{\text{Rayleight-Ritz}}{\leqslant} \rho\left(\left[\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\right]\left[\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\right]\right) \boldsymbol{x}^T_{\mathrm{Im}(\boldsymbol{A})}\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} \\
&\underset{\text{Proposição 1.2.1}}{=} \rho\left(\left[\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\right]\left[\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\right]\right) \boldsymbol{x}^T_{\mathrm{Im}(\boldsymbol{A})}\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} \\
&= \rho\left(\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\left(\boldsymbol{A}^T\boldsymbol{A}\right)\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\right) \boldsymbol{x}^T_{\mathrm{Im}(\boldsymbol{A})}\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} \\
&= \rho\Big(\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\Big) \boldsymbol{x}^T_{\mathrm{Im}(\boldsymbol{A})}\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})}.
\end{aligned}$$

A tese fica estabelecida observando que $\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A})} = \boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A})}\boldsymbol{x}$ onde $\boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A})}$ é a matriz associada à projecção ortogonal sobre o espaço imagem $\mathrm{Im}(\boldsymbol{A})$. ∎

Proposição 1.2.4 *Se* $\boldsymbol{A} \in \mathcal{M}_{\kappa\times n}(\mathbb{R})$ *então*

$$\|\boldsymbol{A}\boldsymbol{x}\|^2 \leqslant \rho\left(\boldsymbol{A}\boldsymbol{A}^T\right) \left\| \boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A}^T)}\boldsymbol{x} \right\|^2, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$$

onde $\boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A}^T)}$ *é a matriz associada à projecção ortogonal sobre o espaço imagem* $\mathrm{Im}(\boldsymbol{A}^T)$.

Demonstração. Seguindo os mesmos passos da Proposição 1.2.3 tem-se $\boldsymbol{Ax} = \boldsymbol{Ax}_{\mathrm{Im}(\boldsymbol{A}^T)}$ uma vez que $\mathbb{R}^\kappa = \mathrm{Nuc}(\boldsymbol{A}) \oplus \mathrm{Im}(\boldsymbol{A}^T)$. Por outro lado, $\boldsymbol{A}^T\boldsymbol{A}$ é simétrica definida não-negativa donde todos os seus valores próprios são não-negativos vindo,

$$\begin{aligned}\|\boldsymbol{Ax}\|^2 = \left\|\boldsymbol{Ax}_{\mathrm{Im}(\boldsymbol{A}^T)}\right\|^2 &= \boldsymbol{x}^T_{\mathrm{Im}(\boldsymbol{A}^T)}\boldsymbol{A}^T\boldsymbol{Ax}_{\mathrm{Im}(\boldsymbol{A}^T)} \leqslant \\ &\underset{\text{Rayleight-Ritz}}{\leqslant} \rho\left(\boldsymbol{A}^T\boldsymbol{A}\right)\boldsymbol{x}^T_{\mathrm{Im}(\boldsymbol{A}^T)}\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A}^T)} \underset{\text{Proposição 1.2.2}}{=} \rho\left(\boldsymbol{AA}^T\right)\left\|\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A}^T)}\right\|^2.\end{aligned}$$

A conclusão segue do facto de $\boldsymbol{x}_{\mathrm{Im}(\boldsymbol{A}^T)} = \boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A}^T)}\boldsymbol{x}$ onde $\boldsymbol{P}_{\mathrm{Im}(\boldsymbol{A}^T)}$ é a matriz associada à projecção ortogonal sobre o espaço imagem $\mathrm{Im}(\boldsymbol{A}^T)$. ■

Uma sucessão de vectores $\{\boldsymbol{x}_m,\, m \geqslant 1\}$ de $\mathbb{R}^n$ converge para um vector $\boldsymbol{x}$ de $\mathbb{R}^n$ se,

$$\lim_{m\to\infty}\|\boldsymbol{v}_m - \boldsymbol{v}\| = 0$$

e indicamos $\lim\limits_{m\to\infty}\boldsymbol{x}_m = \boldsymbol{x}$ ou $\boldsymbol{x}_m \longrightarrow \boldsymbol{x}$ quando $m \to \infty$. Dada uma matriz $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ definimos a norma matricial de $\boldsymbol{A}$ por,

$$\|\boldsymbol{A}\| = \sup_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|\boldsymbol{Ax}\|}{\|\boldsymbol{x}\|}$$

tendo-se então $\|\boldsymbol{A}\| = \sqrt{\rho(\boldsymbol{A}^T\boldsymbol{A})}$; se $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ for simétrica tem-se $\|\boldsymbol{A}\| = \rho(\boldsymbol{A})$. Diz-se que uma sucessão de matrizes $\{\boldsymbol{A}_m,\, m \geqslant 1\}$ de $\mathcal{M}_{\kappa\times n}(\mathbb{R})$ converge para uma matriz $\boldsymbol{A} \in \mathcal{M}_{\kappa\times n}(\mathbb{R})$ se,

$$\lim_{m\to\infty}\|\boldsymbol{A}_m - \boldsymbol{A}\| = 0$$

e escrevemos $\lim\limits_{n\to\infty}\boldsymbol{A}_m = \boldsymbol{A}$ ou $\boldsymbol{A}_m \longrightarrow \boldsymbol{A}$ quando $m \to \infty$.

Apresentamos agora uma versão matricial de uma ferramenta essencial em teoria de probabilidades: o lema de Kronecker.

Lema 1.2.1 (lema de Kronecker matricial) *Seja $\{\boldsymbol{v}_m,\, m \geqslant 1\}$ uma sucessão de vectores de $\mathbb{R}^n$ e $\{\boldsymbol{A}_m,\, m \geqslant 1\}$ uma sucessão de matrizes de $\mathcal{M}_{n\times n}(\mathbb{R})$ simétricas, invertíveis, definidas positivas e tais que $\boldsymbol{A}_{m+1} - \boldsymbol{A}_m$ é definida não-negativa para todo o $m \geqslant 1$. Se,*

(a) $\sum\limits_{i=m}^{\infty}\boldsymbol{A}_i^{-1}\boldsymbol{v}_i$ *existe finito.*

(b) $\lim\limits_{m\to\infty}\boldsymbol{A}_m^{-1} = \mathbf{O}$.

(c) $\limsup\limits_{m\to\infty}\sum\limits_{i=1}^{m}\left\|\boldsymbol{A}_m^{-1}(\boldsymbol{A}_i - \boldsymbol{A}_{i-1})\right\| < \infty$.

então

$$\lim_{m\to\infty}\boldsymbol{A}_m^{-1}\sum_{i=1}^{m}\boldsymbol{v}_i = \mathbf{0}.$$

Demonstração. A demonstração pode ser vista na Secção 2 de [5]. ■

Observação 1.2.3 Designando por $\lambda_{\max}(\boldsymbol{A}_m)$ e $\lambda_{\min}(\boldsymbol{A}_m)$, respectivamente, o maior e o menor valor próprio da matriz $\boldsymbol{A}_m \in \mathcal{M}_{n\times n}(\mathbb{R})$ referenciada no lema de Kronecker matricial então a condição

$$\limsup_{m\to\infty} \frac{\lambda_{\max}(\boldsymbol{A}_m)}{\lambda_{\min}(\boldsymbol{A}_m)} < \infty \tag{1.2.5}$$

implica a hipótese $\limsup_{m\to\infty} \sum_{i=1}^{m} \left|\left|\left| \boldsymbol{A}_m^{-1}(\boldsymbol{A}_i - \boldsymbol{A}_{i-1}) \right|\right|\right| < \infty$. Convém destacar que a conclusão do lema de Kronecker matricial pode ser falsa se a condição (1.2.5) não se verificar (ver [6] para a construção de um contra-exemplo).

Observação 1.2.4 De referir ainda que é possível de estabelecer $\lim_{m\to\infty} \boldsymbol{A}_m^{-1} \sum_{i=1}^{m} \boldsymbol{v}_i = \boldsymbol{0}$ sob hipóteses ligeiramente mais fracas do que as assumidas no Lema 1.2.1 (ver [35] para mais detalhes).

De seguida, abordamos alguns conceitos elementares sobre a distribuição normal multivariada de dimensão $n \geqslant 2$. Recordemos que, um vector aleatório $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ tem distribuição normal multivariada não-degenerada se a sua função característica for,

$$\varphi(\boldsymbol{t}) = \exp\left\{ i\, \boldsymbol{t}^T \boldsymbol{\eta} - \frac{1}{2}\, \boldsymbol{t}^T \boldsymbol{C} \boldsymbol{t} \right\}, \quad \boldsymbol{t} \in \mathbb{R}^n$$

com $\mathrm{car}(\boldsymbol{C}) = n$ (ver Apêndice).

Teorema 1.2.8 *Se o vector aleatório* $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ *tem distribuição normal multivariada não-degenerada com matriz de covariância* $\boldsymbol{C} = \left[\nu_{ij}\right]_{\substack{i=1,\ldots,n\\ j=1,\ldots,n}}$ *então as componentes* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *são independentes se e só se* $\nu_{ij} = 0$ *para todo o* $i \neq j$ $(i, j = 1, \ldots, n)$.

Demonstração. Consultar [81], página 247. ∎

Teorema 1.2.9 *Se o vector aleatório* $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ *tem distribuição normal multivariada não-degenerada com vector médio* $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$ *e matriz de covariância* $\boldsymbol{C} = \mathrm{diag}(\nu_1^2, \ldots, \nu_n^2)$ *então cada componente* X_j *tem distribuição* $N(\eta_j, \nu_j^2)$.

Demonstração. Para cada $j = 1, \ldots, n$ a função característica da componente X_j é,

$$\begin{aligned}\varphi_{\mathrm{X}_j}(t_j) = \mathbb{E}\left(e^{it_j\mathrm{X}_j}\right) = \mathbb{E}\left(e^{i0\mathrm{X}_1 + \ldots + i0\mathrm{X}_{j-1} + it_j\mathrm{X}_j + i0\mathrm{X}_{j+1} + \ldots + i0\mathrm{X}_n}\right) = \\ = \varphi_{\mathbf{X}}(0, \ldots, 0, t_j, 0, \ldots, 0) = e^{it_j\eta_i - \frac{1}{2}\nu_j^2 t_j^2}\end{aligned}$$

e portanto X_j tem distribuição $N(\eta_j, \nu_j^2)$. ∎

As distribuições normais têm um papel crucial em teoria de probabilidades, nomeadamente, nos teoremas do limite central. Muitas variáveis, pelo menos numa primeira aproximação, podem ser consideradas como normais. É o que acontece com os erros de medida em Física, o rendimento de um terreno em Agronomia ou a amplitude de um sinal captado numa linha de ruído,... cujas variações, por vezes muito grandes, devem-se a um vasto número de factores. Finalizamos esta secção com um resultado de enorme importância para observações que se regem segundo uma distribuição normal: o teorema de Cochran.

Teorema 1.2.10 (teorema de Cochran) *Sejam* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *v.a.'s independentes com distribuição normal* $N(0, \nu^2)$ *e* $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$.

(a) *As componentes de* $\mathbf{X}$ *em toda a base ortonormada de* $\mathbb{R}^n$ *são independentes tendo cada uma distribuição* $N(0, \nu^2)$. *Além disso, qualquer transformação ortogonal de* $\mathbf{X}$ *é um vector de* n *componentes independentes onde cada componente tem distribuição* $N(0, \nu^2)$.

(b) *Se* $\mathbb{F}_1 \oplus \ldots \oplus \mathbb{F}_\kappa$ *for uma decomposição de* $\mathbb{R}^n$ *em* κ *subespaços de* $\mathbb{R}^n$ *de dimensões* $m_1, \ldots, m_\kappa$ *então os vectores aleatórios* $\boldsymbol{P}_{\mathbb{F}_1}\mathbf{X}, \ldots, \boldsymbol{P}_{\mathbb{F}_m}\mathbf{X}$ *(respectivamente, as projecções ortogonais de* $\mathbf{X}$ *sobre* $\mathbb{F}_1, \ldots, \mathbb{F}_\kappa$*) são independentes. Adicionalmente, as v.a.'s* $\|\boldsymbol{P}_{\mathbb{F}_1}\mathbf{X}\|^2, \ldots, \|\boldsymbol{P}_{\mathbb{F}_m}\mathbf{X}\|^2$ *são independentes, respectivamente, com distribuição* $\nu^2\chi^2(m_1), \ldots, \nu^2\chi^2(m_\kappa)$.

Demonstração. Consultar [21], página 119. ∎

Sejam $\mathrm{X}_1, \ldots, \mathrm{X}_n$ v.a.'s independentes em que cada X_i tem distribuição $N(\eta_i, \nu^2)$, $i = 1, \ldots, n$. Considerando o vector aleatório $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ e vector real $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$ podemos escrever $\mathbf{X} = \boldsymbol{\eta} + \mathbf{Y}$ com $\mathbf{Y} = (\mathrm{Y}_1, \ldots, \mathrm{Y}_n)$ onde $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$ são independentes e cada Y_i tem distribuição $N(0, \nu^2)$, $i = 1, \ldots, n$.[17] Nestas condições, o teorema de Cochran admite ainda a seguinte generalização.

Corolário 1.2.1 *Sejam* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *v.a.'s independentes onde cada* X_i *tem distribuição normal* $N(\eta_i, \nu^2)$, $i = 1, \ldots, n$ *e* $\mathbb{F}_1 \oplus \ldots \oplus \mathbb{F}_\kappa$ *uma decomposição de* $\mathbb{R}^n$ *em* κ *subespaços de* $\mathbb{R}^n$ *de dimensões* $m_1, \ldots, m_\kappa$. *Se* $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ *então os vectores aleatórios* $\boldsymbol{P}_{\mathbb{F}_1}\mathbf{X}, \ldots, \boldsymbol{P}_{\mathbb{F}_m}\mathbf{X}$ *(respectivamente, as projecções ortogonais de* $\mathbf{X}$ *sobre* $\mathbb{F}_1, \ldots, \mathbb{F}_\kappa$*) são independentes. Além disso, as v.a.'s* $\|\boldsymbol{P}_{\mathbb{F}_1}\mathbf{X}\|^2, \ldots, \|\boldsymbol{P}_{\mathbb{F}_m}\mathbf{X}\|^2$ *são independentes, respectivamente, com distribuição* $\nu^2\chi^2\left(m_1, \|\boldsymbol{P}_{\mathbb{F}_1}\boldsymbol{\eta}\|^2\right)$, $\ldots$, $\nu^2\chi^2\left(m_\kappa, \|\boldsymbol{P}_{\mathbb{F}_\kappa}\boldsymbol{\eta}\|^2\right)$.

Demonstração. Consultar [21], página 121. ∎

1.3 Tópicos de análise real.

O objectivo central desta secção é, por um lado, introduzir alguns conceitos e notações clássicas da análise real relevantes em secções e capítulos ulteriores e, por outro, anunciar o *método dos multiplicadores de Lagrange* essencial em resultados auxiliares para o terceiro capítulo. Abrimos esta secção com a seguinte definição (ver [3], página 264),

Definição 1.3.1 Chamam-se **números de Bernoulli** à sucessão $\{\mathfrak{b}_n\}$ de números racionais dados pela identidade,

$$\frac{x}{e^x - 1} \equiv \sum_{n=0}^{\infty} \frac{\mathfrak{b}_n\, x^n}{n!}, \quad |x| < 2\pi$$

[17] O vector aleatório $\mathbf{Y}$ é conhecido em teoria do Sinal por *ruído branco* e o vector real $\boldsymbol{\eta}$ representa o *sinal emitido*.

Alternativamente, cada número de Bernoulli $\mathfrak{b}_n$ pode também ser definido por,

$$\mathfrak{b}_n = \frac{n!}{2\pi i} \oint_{\mathcal{C}} \frac{z}{e^z - 1} \frac{dz}{z^{n+1}}$$

onde o contorno $\mathcal{C}$ inclui a origem, tem raio inferior a 2π (para evitar os pólos $\pm 2\pi$) e é orientado no sentido anti-horário (ver [8], página 413). Os primeiros números de Bernoulli são: $\mathfrak{b}_0 = 1$, $\mathfrak{b}_1 = -\frac{1}{2}$, $\mathfrak{b}_2 = \frac{1}{6}, \ldots$. Se $n, p \in \mathbb{N}$ e $\{\mathfrak{b}_m\}$ for a sucessão de números de Bernoulli então,

$$\sum_{j=1}^{n} j^p = \frac{1}{p+1} \sum_{i=1}^{p+1} (-1)^{\delta_{ip}} \binom{p+1}{i} \mathfrak{b}_{p+1-i}\, n^i$$

onde $\delta_{ip} = 0$ se $i \neq p$ e $\delta_{ip} = 1$ se $i = p$ designa-se por **fórmula de Faulhaber** (ver [20], página 106).

Se $\{x_n\}$ e $\{a_n\}$ forem duas sucessões de números reais chamamos **transformação de Abel** a,

$$\sum_{j=1}^{n} x_j a_j = x_n A_n + \sum_{j=1}^{n} (x_j - x_{j+1}) A_j$$

onde $A_j = a_1 + \ldots + a_j$ $(j = 1, \ldots, n)$.

Proposição 1.3.1 (lema de Cesàro) *Se $\{a_n\}$ é uma sucessão de números reais positiva, crescente com $\lim\limits_{n\to\infty} a_n = \infty$ e $\{x_n\}$ uma sucessão de números reais tal que $x_n \longrightarrow x_\infty \in \mathbb{R}$ quando $n \to \infty$ então,*

$$\frac{1}{a_n} \sum_{i=1}^{n} (a_i - a_{i-1}) x_i \longrightarrow x_\infty$$

quando $n \to \infty$.

Demonstração. Consultar [91], página 116. ∎

A proposição que se segue é um resultado *sine qua non* em teoria de probabilidades.

Proposição 1.3.2 (lema de Kronecker) *Se $\{a_n\}$ é uma sucessão de números reais positiva, crescente com $\lim\limits_{n\to\infty} a_n = \infty$ e $\{x_n\}$ uma sucessão de números reais tal que $\sum\limits_{n=1}^{\infty} \frac{x_n}{a_n}$ é convergente então,*

$$\frac{1}{a_n} \sum_{i=1}^{n} x_i \longrightarrow 0$$

quando $n \to \infty$.

Demonstração. Consultar [91], página 117. ∎

Proposição 1.3.3 *Se $a_1, \ldots, a_n$ são reais não-negativos então,*

$$\left(a_1^s + \ldots + a_n^s\right)^{1/s} \leqslant \left(a_1^r + \ldots + a_n^r\right)^{1/r}$$

para $0 < r < s$.

Demonstração. Com efeito,

$$\frac{\left(\sum_{i=1}^n a_i^s\right)^{1/s}}{\left(\sum_{j=1}^n a_j^r\right)^{1/r}} = \left[\sum_{i=1}^n \frac{a_i^s}{\left(\sum_{j=1}^n a_j^r\right)^{s/r}}\right]^{1/s} = \left[\sum_{i=1}^n \left(\frac{a_i^r}{\sum_{j=1}^n a_j^r}\right)^{s/r}\right]^{1/s} \leqslant \left(\sum_{i=1}^n \frac{a_i^r}{\sum_{j=1}^n a_j^r}\right)^{1/r} = 1$$

atendendo ao facto de $(\sum_{i=1}^n a_i)^p \geqslant \sum_{i=1}^n a_i^p$, $p \geqslant 1$. ■

Expomos, de seguida, as conhecidas notações de Landau. Sejam $f, g \colon D \subset \mathbb{R} \longrightarrow \mathbb{R}$ funções e $x_0 \in \overline{D}$.

(L1) Se existir uma vizinhança $\mathcal{V}_\varepsilon(x_0)$ de x_0 e uma constante $c > 0$ tal que para todo o $x \in \mathcal{V}_\varepsilon(x_0) \cap D$,

$$|f(x)| \leqslant c\,|g(x)|$$

dizemos que f é **fracamente inferior** a g em x_0 e escreve-se $f = O(g)$, $x \to x_0$. Se as funções f e g são tais que $f = O(g)$ e $g = O(f)$ quando $x \to x_0$ então dizem-se do **mesmo grau** quando $x \to x_0$ e escreve-se $f \asymp g$, $x \to x_0$.

(L2) Se existir uma vizinhança $\mathcal{V}_\varepsilon(x_0)$ de x_0 e uma função $\delta \colon D \subset \mathbb{R} \longrightarrow \mathbb{R}$ verificando $\lim_{x \to x_0} \delta(x) = 0$ tais que para todo o $x \in \mathcal{V}_\varepsilon(x_0) \cap D$,

$$f(x) = \delta(x) g(x)$$

dizemos que f é **fortemente inferior** a g em x_0 e escrevemos $f = o(g)$, $x \to x_0$.

(L3) Se existir uma vizinhança $\mathcal{V}_\varepsilon(x_0)$ de x_0 e uma função $\xi \colon \mathcal{V}_\varepsilon(x_0) \cap D \longrightarrow \mathbb{R}$ verificando $\lim_{x \to x_0} \xi(x) = 1$ tais que para todo o $x \in \mathcal{V}_\varepsilon(x_0) \cap D$,

$$f(x) = \xi(x) g(x)$$

dizemos que f e g são (assimptoticamente) **equivalentes** em x_0 e escrevemos $f \sim g$, $x \to x_0$.

Observação 1.3.1 Dadas as funções $f, g \colon D \subset \mathbb{R} \longrightarrow \mathbb{R}$ e $x_0 \in \overline{D}$ então podemos afirmar que f é fracamente inferior a g em x_0 se e só se $\limsup_{n \to x_0} \left|\frac{f(x)}{g(x)}\right| < \infty$.

Introduzimos agora alguns conceitos adicionais, nomeadamente, o de variação lenta, variação regular e variação rápida de uma função no infinito. Seguiremos de perto o excepcional *Regular variation* de N. H. Bingham, C. M. Goldie e J. L. Teugels onde as definições e as demonstrações dos teoremas aqui anunciados podem ser encontradas. Em *grosso modo* as funções de variação regular são aquelas que, assimptoticamente, se comportam como uma potência x^α, $\alpha \in \mathbb{R}$.

Definição 1.3.2 Seja f uma função positiva e mensurável à Lebesgue em $]0, \infty[$. Então, (R1) f é de **variação lenta** no infinito e escrevemos $f \in \mathcal{R}_0$ se,

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = 1, \quad t > 0.$$

(R2) f é de **variação regular** no infinito de índice $\alpha \in \mathbb{R}$ e escrevemos $f \in \mathcal{R}_\alpha$ se,

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^\alpha, \quad t > 0.$$

(R3) f é de **variação rápida** no infinito e escrevemos $f \in \mathcal{R}_{-\infty}$ se,

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = \begin{cases} \infty & \text{se} \quad 0 < t < 1 \\ 1 & \text{se} \quad t = 1 \\ 0 & \text{se} \quad t > 1 \end{cases}.$$

Teorema 1.3.1 *Se $f \in \mathcal{R}_\alpha$ com $\alpha > 0$ então existe $g \in \mathcal{R}_{1/\alpha}$ tal que*

$$f\big(g(x)\big) \sim x \sim g\big(f(x)\big), \quad x \to \infty.$$

Demonstração. Consultar [11], página 28. ∎

A função g do Teorema 1.3.1 chama-se uma **inversa assimptótica** de f. Se f é uma função definida em $[x_0, \infty[$, localmente limitada e verifica $\lim\limits_{x\to\infty} f(x) = \infty$ então uma versão da sua inversa assimptótica é a **função inversa generalizada** de f definida em $[f(x_0), \infty[$ e dada por,

$$f^{-1}(t) = \inf\big\{x \in [x_0, \infty[:\ f(x) > t\big\}.$$

Teorema 1.3.2 *Se ℓ é uma função de variação lenta então existe uma função de variação lenta $\ell^{\#}$, única a menos de uma equivalência assimptótica, tal que*

$$\lim_{x\to\infty} \ell(x)\, \ell^{\#}\big(x\ell(x)\big) = 1, \qquad \lim_{x\to\infty} \ell^{\#}(x)\, \ell\big(x\ell^{\#}(x)\big) = 1$$

e consequentemente $\ell^{\#\#} \sim \ell$.

Demonstração. Consultar [11], página 29. ∎

A função de variação lenta $\ell^{\#}$ do Teorema 1.3.2 é chamada **conjugada de Bruijn** de ℓ e $(\ell, \ell^{\#})$ diz-se um **par conjugado de Bruijn**.

Teorema 1.3.3 *Sejam $a, b > 0$ e $f(x) \sim x^{ab}\ell^a(x^b)$, $x \to \infty$ onde ℓ é uma função de variação lenta. Se g for uma inversa assimptótica de f então*

$$g(x) \sim x^{\frac{1}{ab}} \left[\ell^{\#}\left(x^{\frac{1}{a}}\right)\right]^{\frac{1}{b}}, \quad x \to \infty.$$

Demonstração. Consultar [11], página 29. ∎

Em particular, se f for uma função definida em $[x_0, \infty[$, localmente limitada e tal que $\lim_{x\to\infty} f(x) = \infty$ então tem-se $f^{-1}(x) \sim x^{\frac{1}{ab}} \left[\ell^{\#}\left(x^{\frac{1}{a}}\right)\right]^{\frac{1}{b}}$ quando $x \to \infty$.

Teorema 1.3.4 *Se $h \in \mathcal{R}_{-\infty}$ é não-crescente então para algum $u > 0$ e todo o $\alpha \in \mathbb{R}$,*

$$\int_u^\infty t^\alpha h(t)\,dt < \infty \qquad e \qquad \lim_{x\to\infty} \frac{x^{\alpha+1}h(x)}{\int_x^\infty t^\alpha h(t)\,dt} = \infty.$$

Reciprocamente, se para algum $\alpha \in \mathbb{R}$, $\int_1^\infty t^\alpha h(t)\,dt < \infty$ e $\lim_{x\to\infty} \frac{x^{\alpha+1}h(x)}{\int_x^\infty t^\alpha h(t)\,dt} = \infty$ então $h \in \mathcal{R}_{-\infty}$.

Demonstração. Consultar [26], página 570. ∎

A passagem para **coordenadas polares** em $\mathbb{R}^n$ está definida pela transformação de $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ em $(r, \theta_1, \ldots, \theta_{n-1}) \in]0,\infty[\times]0,\pi[\times\ldots\times]0,\pi[\times]0,2\pi[$ dada pelas fórmulas,

$$\begin{cases} x_1 = r\cos\theta_1 \\ x_2 = r\sin\theta_1\cos\theta_2 \\ \qquad\vdots \\ x_{n-1} = r\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-2}\cos\theta_{n-1} \\ x_n = r\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-1} \end{cases}.$$

Tem-se ainda $d\boldsymbol{x} = r^{n-1}(\sin\theta_1)^{n-2}(\sin\theta_2)^{n-3}\ldots\sin\theta_{n-2}\,dr\,d\theta_1\ldots d\theta_{n-1}$ e supondo um ponto $\boldsymbol{s} = (s_1, \ldots, s_n)$ da esfera unitária $\mathbb{S}^{n-1}$ podemos escrever para qualquer função $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ que seja $\mathcal{B}(\mathbb{R}^n)$-mensurável e integrável (considerando o triplo $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathrm{Leb})$ onde Leb é a medida de Lebesgue[18]),

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\,d\boldsymbol{x} = \int_0^\infty \int_{\mathbb{S}^{n-1}} f(r\,\boldsymbol{s})r^{n-1}\,dr\,d\boldsymbol{s}.$$

Relembramos, de seguida, a noção de ponto de máximo e mínimo de uma função real de várias variáveis.

Definição 1.3.3 Suponhamos a função $f\colon D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$. Um ponto $\boldsymbol{x}_0 \in D$ diz-se um **ponto de máximo** (resp. **ponto de mínimo**) se existe uma bola aberta $\mathcal{B}_\varepsilon(\boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathbb{R}^n\colon \|\boldsymbol{x} - \boldsymbol{x}_0\| < \varepsilon\}$ tal que para todo $\boldsymbol{x} \in \mathcal{B}_\varepsilon(\boldsymbol{x}_0) \cap D$ se tenha a desigualdade $f(\boldsymbol{x}) \leqslant f(\boldsymbol{x}_0)$ (resp. $f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}_0)$).

Os pontos de máximo e mínimo de uma função designam-se por **pontos de extremo**.

Consideremos agora um aberto $U \subset \mathbb{R}^n$, funções $g_1, \ldots, g_m\colon U \longrightarrow \mathbb{R}$ e o conjunto,

$$\mathcal{K} = \left\{\boldsymbol{x} \in U\colon\ g_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m\right\}$$

dos pontos de U onde as funções $g_1, \ldots, g_m$ se anulam. As equações $g_i(\boldsymbol{x}) = 0$ $(i = 1, \ldots, m)$ chamam-se **ligações**.

[18] A medida de Lebesgue sobre $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ é a medida produto de n exemplares da medida de Lebesgue sobre os borelianos de $\mathbb{R}$ (ver Exemplo 1.1.4).

Definição 1.3.4 Consideremos as funções $f, g_1, \ldots, g_m \colon U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ e o conjunto,

$$\mathcal{K} = \big\{\boldsymbol{x} \in U\colon\ g_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m\big\}.$$

O ponto $\boldsymbol{x}_0 \in \mathcal{K}$ chama-se **extremo condicionado** da função f relativamente às ligações $g_i(\boldsymbol{x}) = 0$ $(i = 1, \ldots, m)$ se $\boldsymbol{x}_0$ for um ponto de extremo da função f restringida ao conjunto $\mathcal{K}$.

Teorema 1.3.5 (multiplicadores de Lagrange) *Seja $U \subset \mathbb{R}^n$ um aberto. Se*

(a) *$f, g_1, \ldots, g_m \colon U \longrightarrow \mathbb{R}$ são funções continuamente deriváveis.*

(b) *$\boldsymbol{x}_0 \in U$ é um extremo condicionado da função f relativamente às ligações $g_i(\boldsymbol{x}) = 0$ $(i = 1, \ldots, m)$.*

então os vectores $\operatorname{grad}\big(f(\boldsymbol{x}_0)\big), \operatorname{grad}\big(g_1(\boldsymbol{x}_0)\big), \ldots, \operatorname{grad}\big(g_m(\boldsymbol{x}_0)\big)$ são linearmente dependentes, isto é, existem números $\nu, \nu_1, \ldots, \nu_m$ tais que,

$$\nu \operatorname{grad}\big(f(\boldsymbol{x}_0)\big) + \nu_1 \operatorname{grad}\big(g_1(\boldsymbol{x}_0)\big) + \ldots + \nu_m \operatorname{grad}\big(g_m(\boldsymbol{x}_0)\big) = \mathbf{0}.$$

Demonstração. Consultar [51], página 101. ∎

Corolário 1.3.1 *Seja $U \subset \mathbb{R}^n$ um aberto. Se*

(a) *$f, g_1, \ldots, g_m \colon U \longrightarrow \mathbb{R}$ são funções continuamente deriváveis.*

(b) *$\boldsymbol{x}_0 \in U$ é um extremo condicionado da função f relativamente às ligações $g_i(\boldsymbol{x}) = 0$ $(i = 1, \ldots, m)$.*

(c) *A característica da matriz jacobiana $\left[\dfrac{\partial g_j}{\partial x_i}(\boldsymbol{x}_0)\right]_{\substack{j=1,\ldots,m\\ i=1,\ldots,n}}$ é m.*

então existem reais $\nu_1, \ldots, \nu_m$ tais que,

$$\operatorname{grad}\big(f(\boldsymbol{x}_0)\big) + \sum_{j=1}^{m} \nu_j \operatorname{grad}\big(g_j(\boldsymbol{x}_0)\big) = 0.$$

Demonstração. Consultar [51], página 101. ∎

A função $L(\boldsymbol{x}) = f(\boldsymbol{x}) + \nu_1 g_1(\boldsymbol{x}) + \ldots + \nu_m g_m(\boldsymbol{x})$ designa-se por **função de Lagrange** e os números reais $\nu_1, \ldots, \nu_m$ chamam-se **multiplicadores de Lagrange**.

As proposições que se seguem, resultam da aplicação do método dos multiplicadores de Lagrange e são de crucial importância para o terceiro capítulo.

Proposição 1.3.4 *Dado o conjunto,*

$$\mathbb{H} = \big\{(t_1, \ldots, t_n) \in \mathbb{R}^n\colon\ t_1 \geqslant 0, \ldots, t_n \geqslant 0,\ t_1^2 + \ldots + t_n^2 = 1\big\}$$

e uma qualquer sucessão $\{a_n\}$ positiva tem-se,

(a) $\max\limits_{\mathbb{H}} \sum\limits_{j=1}^{n} \big(a_j t_j\big)^2 = \left(\max\limits_{1\leqslant j\leqslant n} a_j\right)^2$

(b) $\max\limits_{\mathbb{H}} \sum\limits_{j=1}^{n} \big(a_j t_j\big)^\alpha = \left(a_1^{\frac{2\alpha}{2-\alpha}} + \ldots + a_n^{\frac{2\alpha}{2-\alpha}}\right)^{1-\frac{\alpha}{2}} \qquad (0<\alpha<2)$

Demonstração.

(a) A função $f(t_1,\ldots,t_n) = \big(a_1 t_1\big)^2 + \ldots + \big(a_n t_n\big)^2$ tem um máximo na hiperfície fechada e limitada $\mathbb{H}$ pelo teorema de Weierstrass. Escrevendo $a_m^2 = \max\limits_{1\leqslant j\leqslant n} a_j^2$ é fácil verificar que,

$$f(t_1,\ldots,t_n) = a_m^2 + \sum_{\substack{j=1\\ j\neq m}}^{n} \left(a_j^2 - a_m^2\right) t_j^2 \tag{1.3.1}$$

para todo o $(t_1,\ldots,t_n)\in\mathbb{H}$. Então o máximo de (1.3.1) é atingindo para $t_j = 0$ para todo o $j = 1,\ldots,n$ com $j\neq m$ e $t_m = 1$ implicando,

$$\max_{\mathbb{H}} \sum_{j=1}^{n} \big(a_j t_j\big)^2 = \max_{1\leqslant j\leqslant n} a_j^2 = \left(\max_{1\leqslant j\leqslant n} a_j\right)^2.$$

(b) Pelo teorema de Weierstrass a função

$$f(t_1,\ldots,t_n) = \big(a_1 t_1\big)^\alpha + \ldots + \big(a_n t_n\big)^\alpha, \qquad 0<\alpha<2$$

tem um máximo na hiperfície fechada e limitada $\mathbb{H}$. Se f atingir o seu valor máximo f^* nalgum ponto interno da hiperfície $\mathbb{H}$ então podemos utilizar o método dos multiplicadores de Lagrange para o encontrar. Considerando a ligação,

$$g(t_1,\ldots,t_n) = t_1^2 + \ldots + t_n^2 = 1. \tag{1.3.2}$$

e a função de Lagrange $L(t_1,\ldots,t_n) = f(t_1,\ldots,t_n) + \nu g(t_1,\ldots,t_n)$ tem-se,

$$\begin{cases} \dfrac{\partial L}{\partial t_1} = \alpha a_1^\alpha t_1^{\alpha-1} + 2\nu t_1 = 0 \\ \vdots \\ \dfrac{\partial L}{\partial t_n} = \alpha a_n^\alpha t_n^{\alpha-1} + 2\nu t_n = 0 \end{cases}$$

Deste modo, o multiplicador de Lagrange é $\nu = -\dfrac{\alpha}{2}\left(a_1^{\frac{2\alpha}{2-\alpha}} + \ldots + a_n^{\frac{2\alpha}{2-\alpha}}\right)^{\frac{2-\alpha}{2}}$ e

$$(t_1,\ldots,t_n) = \left(\frac{a_1^{\frac{\alpha}{2-\alpha}}}{\sqrt{a_1^{\frac{2\alpha}{2-\alpha}} + \ldots + a_n^{\frac{2\alpha}{2-\alpha}}}},\ldots,\frac{a_n^{\frac{\alpha}{2-\alpha}}}{\sqrt{a_1^{\frac{2\alpha}{2-\alpha}} + \ldots + a_n^{\frac{2\alpha}{2-\alpha}}}}\right)$$

donde,

$$\max_{\mathbb{H}} \sum_{j=1}^{n} \big(a_j t_j\big)^\alpha = \left(a_1^{\frac{2\alpha}{2-\alpha}} + \ldots + a_n^{\frac{2\alpha}{2-\alpha}}\right)^{1-\frac{\alpha}{2}}.$$

Vejamos agora que a possibilidade de f atingir o seu valor máximo num ponto da fronteira de $\mathbb{H}$ está completamente excluída. Sem perda de generalidade, suponhamos que f atinge o seu valor máximo num ponto da fronteira de $\mathbb{H}$ com a última componente t_n nula. Então, podemos usar novamente o método dos multiplicadores de Lagrange para achar o ponto de extremo no conjunto,

$$\mathbb{H}' = \left\{(t_1, \ldots, t_{n-1}) \in \mathbb{R}^{n-1} \colon\ t_1 \geqslant 0, \ldots, t_{n-1} \geqslant 0,\ \ t_1^2 + \ldots + t_{n-1}^2 = 1\right\}.$$

Cálculos idênticos aos efectuados em cima produzem,

$$\max_{\mathbb{H}'} \sum_{j=1}^{n-1} \left(a_j t_j\right)^{\alpha} = \left(a_1^{\frac{2\alpha}{2-\alpha}} + \ldots + a_{n-1}^{\frac{2\alpha}{2-\alpha}}\right)^{1-\frac{\alpha}{2}} <$$
$$< \left(a_1^{\frac{2\alpha}{2-\alpha}} + \ldots + a_{n-1}^{\frac{2\alpha}{2-\alpha}} + a_n^{\frac{2\alpha}{2-\alpha}}\right)^{1-\frac{\alpha}{2}} = \max_{\mathbb{H}} \sum_{j=1}^{n} \left(a_j t_j\right)^{\alpha}.$$

Naturalmente, se o ponto de extremo na fronteira de $\mathbb{H}$ possuir mais do que uma componente nula, podemos proceder analogamente usando hiperfícies de $\mathbb{R}^{n-p+1}$, em que p é o número de componentes nulas do ponto de extremo. ■

Proposição 1.3.5 *Dado o conjunto,*

$$\mathbb{H} = \left\{(t_1, \ldots, t_n) \in \mathbb{R}^{n} \colon\ t_1 > 0, \ldots, t_n > 0,\ \ t_1^2 + \ldots + t_n^2 = 1\right\}$$

temos,

$$\max_{\mathbb{H}} \sum_{j=1}^{n} t_i \log\left(\frac{1}{t_i}\right) = \frac{1}{2}\sqrt{n}\log n.$$

Além disso, para qualquer sucessão positiva $\{a_n\}$ *tem-se,*

$$\frac{a_1 + \ldots + a_n}{n} \cdot \frac{\sqrt{n}\log n}{2} \leqslant \max_{\mathbb{H}} \sum_{j=1}^{n} a_j t_j \log\left(\frac{1}{t_j}\right) \leqslant \max_{1\leqslant i\leqslant n} a_i \cdot \frac{\sqrt{n}\log n}{2}.$$

Demonstração. Suponhamos a função $f(t_1, \ldots, t_n) = t_1 \log\left(t_1^{-1}\right) + \ldots + t_n \log\left(t_n^{-1}\right)$ e a ligação (1.3.2). A função f admite um máximo na hiperfície $\mathbb{H}$ pois se considerarmos o seu prolongamento por continuidade $\bar{f}$ a $\overline{\mathbb{H}}$ então o teorema de Weierstrass garante que $\bar{f}$ atinge o seu máximo em $\overline{\mathbb{H}}$ e argumentando como na demonstração da Proposição 1.3.4 concluímos que o ponto de extremo terá que ser necessariamente interno de $\overline{\mathbb{H}}$. Deste modo, admitindo a função de Lagrange $L(t_1, \ldots, t_n) = f(t_1, \ldots, t_n)$ $+\ \nu g(t_1, \ldots, t_n)$ vem,

$$\left\{\begin{array}{l} \dfrac{\partial L}{\partial t_1} = -1 + 2\nu t_1 - \log t_1 = 0 \\ \qquad\qquad\vdots \\ \dfrac{\partial L}{\partial t_n} = -1 + 2\nu t_n - \log t_n = 0 \end{array}\right. \iff \left\{\begin{array}{l} \dfrac{1 + \log t_1}{2t_1} = \nu \\ \qquad\vdots \\ \dfrac{1 + \log t_n}{2t_n} = \nu \end{array}\right.$$

para algum ν, independente de j. Visto que $h(t) = \frac{1+\log t}{2t}$ definida em $]0,1]$ é injectiva tem-se $(t_1,\ldots,t_n) = \left(\frac{1}{\sqrt{n}},\ldots,\frac{1}{\sqrt{n}}\right)$ donde,

$$\max_{\mathbb{H}} \sum_{j=1}^{n} t_i \log\left(\frac{1}{t_i}\right) = f\left(\frac{1}{\sqrt{n}},\ldots,\frac{1}{\sqrt{n}}\right) = \frac{1}{2}\sqrt{n}\log n.$$

Para o enquadramento suponhamos,

$$\begin{aligned}
f_1(t_1,\ldots,t_n) &= a_1 t_1 \log\left(\frac{1}{t_1}\right) + a_2 t_2 \log\left(\frac{1}{t_2}\right) + \ldots + a_n t_n \log\left(\frac{1}{t_n}\right)\\
f_2(t_1,\ldots,t_n) &= a_2 t_1 \log\left(\frac{1}{t_1}\right) + a_3 t_2 \log\left(\frac{1}{t_2}\right) + \ldots + a_1 t_n \log\left(\frac{1}{t_n}\right)\\
&\vdots\\
f_n(t_1,\ldots,t_n) &= a_n t_1 \log\left(\frac{1}{t_1}\right) + a_1 t_2 \log\left(\frac{1}{t_2}\right) + \ldots + a_{n-1} t_n \log\left(\frac{1}{t_n}\right).
\end{aligned}$$

Então,

$$n \max_{\mathbb{H}}(f_1) \geqslant \max_{\mathbb{H}}(f_1 + \ldots + f_n)$$

pois $\max_{\mathbb{H}}(f_1) = \max_{\mathbb{H}}(f_2) = \ldots = \max_{\mathbb{H}}(f_n)$ o que implica,

$$\max_{\mathbb{H}} \sum_{j=1}^{n} a_j t_j \log\left(\frac{1}{t_j}\right) \geqslant \frac{a_1 + \ldots + a_n}{n} \cdot \frac{\sqrt{n}\log n}{2}.$$

Por outro lado,

$$\max_{\mathbb{H}} \sum_{j=1}^{n} a_j t_j \log\left(\frac{1}{t_j}\right) \leqslant \max_{1\leqslant i\leqslant n} a_i \cdot \max_{\mathbb{H}} \sum_{j=1}^{n} t_j \log\left(\frac{1}{t_j}\right) = \max_{1\leqslant i\leqslant n} a_i \cdot \frac{\sqrt{n}\log n}{2}$$

como pretendíamos. ∎

Observação 1.3.2 Note-se que se a sucessão $\{a_n\}$ for monótona crescente então,

$$\max_{\mathbb{H}} \sum_{j=1}^{n} a_j t_j \log\left(\frac{1}{t_j}\right) \sim \frac{a_n\sqrt{n}\log n}{2}, \quad n \to \infty$$

pois $\liminf_{n\to\infty} a_n \leqslant \liminf_{n\to\infty} \frac{a_1 + \ldots + a_n}{n}$ (ver [60], página 19).

1.4 Distribuições estáveis unidimensionais

Esta secção é dedicada a uma classe especial de funções de distribuição com grande relevância em capítulos futuros: as distribuições estáveis unidimensionais. Seguiremos de perto o essencial *Stable non-gaussian random processes: stochastic models with infinite variance* de G. Samorodnitsky e M. S. Taqqu; todavia, recomendamos ainda a consulta dos clássicos [29] e [32], bem como de outros livros de texto mais recentes (nomeadamente [7], [41], [89] ou [93]), para uma maior comodidade de leitura dos resultados que se vão expor.

A teoria das distribuições estáveis unidimensionais teve a sua génese por volta de 1920 e 1930 por P. Lévy e A. Y. Khinchin. A grande dispersão das distribuições estáveis é uma das razões da sua enorme importância na modelação dos fenómenos. De facto, qualquer distribuição estável (não-gaussiana) possui grande variabilidade, no sentido em que, é muito mais provável tomar valores distantes do valor médio. As distribuições estáveis têm sido usadas para modelar fenómenos tão diversos como campos gravitacionais de estrelas, distribuições de temperatura de reactores nucleares, pressões em reticulados cristalinos, preços em mercados de capitais, ruído em sistemas de comunicação e ainda de intempéries (aguaceiros anuais).

Nesta secção, abordaremos quatro definições de distribuição estável unidimensional, todas elas equivalentes entre si. As primeiras duas definições são relativas à propriedade "estabilidade": a família de distribuições estáveis é preservada sob convolução.

Definição 1.4.1 Diz-se que uma variável aleatória Z tem uma distribuição **estável** se para quaisquer reais positivos c_1 e c_2 existir um número positivo a e um real b tais que,

$$c_1\,\mathrm{Z}_1 + c_2\,\mathrm{Z}_2 \stackrel{\mathrm{d}}{=} a\,\mathrm{Z} + b \tag{1.4.1}$$

onde Z_1 e Z_2 são cópias[19] independentes de Z. Quando $b = 0$ a variável aleatória Z diz-se **estritamente estável**

Observação 1.4.1 Convém observar que qualquer v.a. X concentrada num único ponto (i.e. degenerada) é sempre estável (basta tomar $a = c_1 + c_2$ e $b = 0$). De notar ainda que qualquer v.a. simétrica e estável é estritamente estável.

Teorema 1.4.1 *Para qualquer v.a.* X *estável existe um número* $\alpha \in]0, 2]$ *tal que o número positivo* a *de* (1.4.1) *satisfaz* $a^\alpha = c_1^\alpha + c_2^\alpha$.

Demonstração. Consultar [82], página 3. ∎

O número α do Teorema 1.4.1 designa-se por **índice de estabilidade** ou **expoente característico**. Uma v.a. X estável com índice de estabilidade α é chamada **$\boldsymbol{\alpha}$-estável**.

Exemplo 1.4.1

Se X for uma v.a. gaussiana com média η e variância ν^2 então X é estável com expoente característico $\alpha = 2$ uma vez que a igualdade em distribuição (1.4.1) é verificada com $a = \sqrt{c_1^2 + c_2^2}$ e $b = \eta\left(c_1 + c_2 - \sqrt{c_1^2 + c_2^2}\right)$.

Definição 1.4.2 Diz-se que uma variável aleatória Z tem uma distribuição **estável** se para todo o $n \geqslant 2$ existir um número positivo a_n e um real b_n tais que,

$$\mathrm{Z}_1 + \ldots + \mathrm{Z}_n \stackrel{\mathrm{d}}{=} a_n \mathrm{Z} + b_n \tag{1.4.2}$$

onde $\mathrm{Z}_1, \ldots, \mathrm{Z}_n$ são cópias independentes de Z.

[19] O termo "cópias"significa que as v.a.'s Z_1 e Z_2 têm a mesma distribuição da v.a. Z. Mais geralmente, diremos que as v.a's $\mathrm{X}_1, \ldots, \mathrm{X}_n$ são cópias da v.a. X quando $\mathrm{X}_1, \ldots, \mathrm{X}_n$ tiverem a mesma distribuição da v.a. X.

Observação 1.4.2 Se Z for uma v.a. α-estável então o número a_n de (1.4.2) é dado necessariamente por $a_n = n^{1/\alpha}$.

A terceira definição de estabilidade está relacionada com o papel das distribuições estáveis no contexto do teorema do limite central e estabelece que as distribuições estáveis são as únicas distribuições que podem ser obtidas como limites de somas de variáveis aleatórias i.i.d. centralizadas e normalizadas. De facto, as distribuições estáveis aproximam a distribuição de somas de v.a.'s i.i.d. centralizadas e normalizadas, tornando-as úteis no modelamento da contribuição de alguns pequenos efeitos casuais.

Definição 1.4.3 Diz-se que uma variável aleatória Z tem uma distribuição **estável** se existir uma sucessão de v.a.'s $\mathrm{Y}_1, \mathrm{Y}_2, \ldots$ i.i.d., uma sucessão de números positivos $\{c_n\}$ e uma sucessão de números reais $\{d_n\}$ tais que,

$$\frac{\mathrm{Y}_1 + \mathrm{Y}_2 + \ldots + \mathrm{Y}_n}{c_n} + d_n \xrightarrow{\mathrm{d}} \mathrm{Z}.$$

Observação 1.4.3 Se Z for gaussiana e as v.a.'s $\mathrm{Y}_1, \mathrm{Y}_2, \ldots$ forem i.i.d. com variância finita então a condição (1.4.3) não é mais do que o bem conhecido teorema do limite central.

Considerando uma sucessão de v.a.'s $\mathrm{X}, \mathrm{X}_1, \mathrm{X}_2, \ldots$ i.i.d. então, se existir uma sucessão de números positivos $\{c_n\}$ e uma sucessão de números reais $\{d'_n\}$ tais que,

$$\frac{\mathrm{X}_1 + \mathrm{X}_2 + \ldots + \mathrm{X}_n - d'_n}{c_n} \xrightarrow{\mathrm{d}} \mathrm{Z} \tag{1.4.3}$$

dizemos que Z (ou a sua função de distribuição) tem um **domínio de atracção** e que a v.a. X pertence ao domínio de atracção de Z. A v.a. X pertence ao **domínio de atracção normal** de uma v.a. α-estável Z quando a sucessão normalizante c_n for dada por $c_n = c\, n^{1/\alpha}$ para alguma constante c positiva.

A última definição especifica a função característica de uma v.a. estável Z. Efectivamente, esta definição costuma ser usada extensivamente e a principal razão, prende-se com o facto da caracterização de uma v.a. Z estável ser feita através de quatro parâmetros (únicos).

Definição 1.4.4 Diz-se que uma variável aleatória Z tem uma distribuição **estável** se existirem parâmetros $0 < \alpha \leqslant 2$, $\sigma \geqslant 0$, $-1 \leqslant \lambda \leqslant 1$ e $\mu \in \mathbb{R}$ tais que a sua função característica assuma a forma,

$$\mathbb{E}\Big(e^{i\mathrm{Z}t}\Big) = \begin{cases} \exp\left(-\sigma^{\alpha}\, |t|^{\alpha} \left(1 - i\lambda \operatorname{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right)\right) + i\mu t\right) & \text{se} \quad \alpha \neq 1 \\ \exp\left(-\sigma\, |t| \left(1 + i\lambda \frac{2}{\pi} \operatorname{sign}(t) \log|t|\right) + i\mu t\right) & \text{se} \quad \alpha = 1 \end{cases} \tag{1.4.4}$$

com $t \in \mathbb{R}$ e onde,

$$\operatorname{sign}(t) = \begin{cases} 1 & \text{se} \quad t > 0 \\ 0 & \text{se} \quad t = 0 \\ -1 & \text{se} \quad t < 0 \end{cases} .$$

Deste modo, a estabilidade de uma variável aleatória é caracterizada pelos quatro parâmetros únicos,

$$\alpha \in]0, 2], \quad \sigma \geqslant 0, \quad \lambda \in [-1, 1], \quad \mu \in \mathbb{R}$$

respectivamente, o **índice de estabilidade**, **parâmetro de escala**, **parâmetro de distorção** e **parâmetro de localização**. Designaremos as distribuições estáveis por $\mathcal{S}_\alpha(\sigma, \lambda, \mu)$ e uma variável aleatória estável Z será indicada escrevendo,

$$\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu).^{20}$$

Já vimos que qualquer v.a. gaussiana com média η e variância ν^2 é estável com expoente característico $\alpha = 2$. Porém, é sabido que a sua função característica tem a forma $\exp\left(-\frac{t^2\nu^2}{2} + i\eta t\right)$, ou seja, $\alpha = 2$, $\sigma = \sqrt{\frac{\nu^2}{2}}$ e $\mu = \eta$ em (1.4.4) donde se conclui que o parâmetro de distorção λ é totalmente irrelevante quando $\alpha = 2$. Na representação acima, uma constante μ (v.a. degenerada) tem distribuição degenerada $\mathcal{S}_\alpha(0, \lambda, \mu)$ para qualquer $0 < \alpha \leqslant 2$. Se nada for dito em contrário, assumiremos doravante que $\sigma > 0$ convencionando, assim, a exclusão das distribuições degeneradas deste texto já que estas possuem propriedades bastante invulgares (por exemplo, todos os momentos de uma distribuição degenerada são finitos mas distribuições α-estáveis não-degeneradas com $0 < \alpha < 2$ têm momentos de segunda ordem infinitos). Uma distribuição $\mathcal{S}_\alpha(\sigma, \lambda, 0)$ diz-se **distorcida à direita** se $\lambda > 0$; se $\lambda < 0$ a distribuição $\mathcal{S}_\alpha(\sigma, \lambda, 0)$ diz-se **distorcida à esquerda**. Dizemos ainda que a distribuição $\mathcal{S}_\alpha(\sigma, 1, 0)$ é **totalmente distorcida à direita** e que a distribuição $\mathcal{S}_\alpha(\sigma, -1, 0)$ é **totalmente distorcida à esquerda**. O suporte de qualquer distribuição $\mathcal{S}_\alpha(\sigma, 1, \mu)$ com $0 < \alpha < 1$ é $[\mu, \infty[$; analogamente, o suporte de qualquer distribuição $\mathcal{S}_\alpha(\sigma, -1, \mu)$ com $0 < \alpha < 1$ é $]-\infty, \mu]$.

Observação 1.4.4 As funções de distribuição de v.a.'s α-estáveis (isto é, v.a.'s estáveis com expoente característico α) não-degeneradas são absolutamente contínuas e as suas densidades infinitamente diferenciáveis em todo o ponto da recta real.

A função característica (1.4.4) é uma ferramenta poderosa no estudo das v.a.'s estáveis, em particular, das suas propriedades elementares.

[20] Observemos que existe uma coincidência com o símbolo "$\sim$"utilizado para indicar que duas funções são equivalentes num ponto x_0 (conf. Secção 1.3). Contudo, não há perigo de confusão com a notação $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$ já que na equivalência de funções expliciteremos sempre o ponto em causa, na circunstância escrevendo "$x \to x_0$".

Proposição 1.4.1 *Se* Z_1 *e* Z_2 *são v.a.'s independentes tais que* $\mathrm{Z}_1 \sim \mathcal{S}_\alpha(\sigma_1, \lambda_1, \mu_1)$ *e* $\mathrm{Z}_2 \sim \mathcal{S}_\alpha(\sigma_2, \lambda_2, \mu_2)$ *então* $\mathrm{Z}_1 + \mathrm{Z}_2 \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$ *com,*

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}, \quad \lambda = \frac{\lambda_1\sigma_1^\alpha + \lambda_2\sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad \mu = \mu_1 + \mu_2.$$

Demonstração. Consultar [82], página 10. ∎

Proposição 1.4.2 *Se* $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$ *e* τ *é uma constante real então* $\mathrm{Z} + \tau \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu + \tau)$.

Demonstração. Consultar [82], página 11. ∎

Proposição 1.4.3 *Se* $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$ *e* τ *é uma constante real não nula então*

$$\begin{aligned} \tau\mathrm{Z} &\sim \mathcal{S}_\alpha(|\tau|\,\sigma, \operatorname{sign}(\tau)\lambda, \tau\mu) && \text{se} \quad \alpha \neq 1 \\ \tau\mathrm{Z} &\sim \mathcal{S}_1\left(|\tau|\,\sigma, \operatorname{sign}(\tau)\lambda, \tau\mu - \frac{2}{\pi}\sigma\lambda\tau\log|\tau|\right) && \text{se} \quad \alpha = 1. \end{aligned}$$

Demonstração. Consultar [82], página 11. ∎

Se $\mathrm{Z} \sim \mathcal{S}_\alpha(1, \lambda, 0)$ então as duas últimas propriedades permitem escrever,

$$\sigma\mathrm{Z} + \mu \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu) \quad \text{se} \quad \alpha \neq 1$$

e

$$\sigma\mathrm{Z} + \frac{2}{\pi}\lambda\sigma\log\sigma + \mu \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu) \quad \text{se} \quad \alpha = 1.$$

Proposição 1.4.4 *Para qualquer* $0 < \alpha < 2$,

$$\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, 0) \quad \Longleftrightarrow \quad -\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, -\lambda, 0).$$

Demonstração. Consultar [82], página 11. ∎

Proposição 1.4.5 $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$ *é simétrica relativamente ao ponto* μ *se e só se* $\lambda = 0$.

Demonstração. Consultar [82], página 11. ∎

Proposição 1.4.6 *Se* $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$ *com* $0 < \alpha \leqslant 2$, $\alpha \neq 1$ *então* Z *é estritamente estável se e só se* $\mu = 0$.

Demonstração. Consultar [82], página 12. ∎

Proposição 1.4.7 $\mathrm{Z} \sim \mathcal{S}_1(\sigma, \lambda, \mu)$ *é estritamente estável se e só se* $\lambda = 0$.

Demonstração. Consultar [82], página 12. ∎

O próximo resultado dá-nos a possibilidade de obter $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, 0)$, $0 < \alpha < 2$ à custa de duas v.a.'s independentes Y_1 e Y_2 ambas com distribuição $\mathcal{S}_\alpha(\sigma, 1, 0)$.

Teorema 1.4.2 *Se* $\mathrm{Z} \sim (\sigma, \lambda, 0)$, $0 < \alpha < 2$ *então existem duas v.a.'s* Y_1 *e* Y_2 *i.i.d. com distribuição comum* $\mathcal{S}_\alpha(\sigma, 1, 0)$ *tais que,*

$$\mathrm{Z} \stackrel{\mathrm{d}}{=} \left(\frac{1+\lambda}{2}\right)^{1/\alpha} \mathrm{Y}_1 - \left(\frac{1-\lambda}{2}\right)^{1/\alpha} \mathrm{Y}_2 \quad \textit{se} \;\; \alpha \neq 1$$

e

$$\mathrm{Z} \stackrel{\mathrm{d}}{=} \left(\frac{1+\lambda}{2}\right) \mathrm{Y}_1 - \left(\frac{1-\lambda}{2}\right) \mathrm{Y}_2 + \sigma \left[\frac{1+\lambda}{\pi} \log\left(\frac{1+\lambda}{2}\right) - \frac{1-\lambda}{\pi} \log\left(\frac{1-\lambda}{2}\right)\right] \quad \textit{se} \;\; \alpha = 1$$

Demonstração. Consultar [82], página 16. ■

Quando temos uma v.a. gaussiana, isto é, quando $\mathrm{Z} \sim \mathcal{S}_2(\sigma, *, \mu)$ conhecemos perfeitamente o comportamento assimptótico das probabilidades de cauda,

$$\mathbb{P}\{\mathrm{Z} < -z\} = \mathbb{P}\{\mathrm{Z} > z\} \sim \frac{1}{2\sqrt{\pi}\sigma z} e^{-\frac{z^2}{4\sigma^2}}, \quad z \to \infty.$$

No entanto, quando $0 < \alpha < 2$ o comportamento assimptótico das probabilidades de cauda é $z^{-\alpha}$.

Teorema 1.4.3 *Se* $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$, $0 < \alpha < 2$ *então,*

$$\lim_{z\to\infty} z^\alpha \mathbb{P}\{\mathrm{Z} > z\} = c(\alpha)\frac{1+\lambda}{2}\sigma^\alpha \quad e \quad \lim_{z\to\infty} z^\alpha \mathbb{P}\{\mathrm{Z} < -z\} = c(\alpha)\frac{1-\lambda}{2}\sigma^\alpha$$

onde

$$c(\alpha) = \left(\int_0^\infty t^{-\alpha} \sin t \, dt\right)^{-1} = \begin{cases} \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos\left(\frac{\alpha\pi}{2}\right)} & \text{se} \quad \alpha \neq 1 \\[2ex] \dfrac{2}{\pi} & \text{se} \quad \alpha = 1 \end{cases}$$

com $\Gamma(\,\cdot\,)$ *a função gama*[21].

Demonstração. Consultar [82], página 16. ■

Os resultados seguintes são referentes aos momentos absolutos de ordem $p > 0$ de uma v.a. estável. Em particular, qualquer v.a. α-estável com $0 < \alpha < 2$ tem variância infinita; inclusive, toda a v.a. α-estável com $0 < \alpha \leqslant 1$ não é integrável, não possuindo assim valor médio.

Teorema 1.4.4 *Se* $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$, $0 < \alpha < 2$ *então,*

$$\mathbb{E}\,|\mathrm{Z}|^p < \infty \;\; \textit{para qualquer} \;\; 0 < p < \alpha,$$
$$\mathbb{E}\,|\mathrm{Z}|^p = \infty \;\; \textit{para qualquer} \;\; p \geqslant \alpha.$$

[21] $\Gamma(x) = \displaystyle\int_0^\infty t^{x-1} e^{-t} \, dt$, $x > 0$.

Demonstração. Consultar [82], página 18. ∎

Corolário 1.4.1 *Se* $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, 0)$ *com* $0 < \alpha < 2$, $\alpha \neq 1$ *então para todo o* $0 < p < \alpha$,

$$\mathbb{E}\,|\mathrm{Z}|^p = \sigma^p \cdot \frac{2^{p-1}\Gamma\left(1-\frac{p}{\alpha}\right)}{p\int_0^\infty t^{-p-1}\sin^2 t\,dt}\left[1+\lambda^2\tan^2\left(\frac{\alpha\pi}{2}\right)\right]^{\frac{p}{2\alpha}} \cdot \cos\left(\frac{p}{\alpha}\,\arctan\left(\lambda\tan\left(\frac{\alpha\pi}{2}\right)\right)\right).$$

Se $\mathrm{Z} \sim \mathcal{S}_1(\sigma, 0, 0)$ *então para todo o* $0 < p < 1$,

$$\mathbb{E}\,|\mathrm{Z}|^p = \frac{\sigma^p\,2^{p+1}\,\Gamma\left(\frac{p+1}{2}\right)\Gamma(-p)}{\sqrt{\pi}\,\Gamma\left(-\frac{p}{2}\right)}.$$

Demonstração. Consultar [82], página 18. ∎

Observação 1.4.5 O momento absoluto de ordem $-1 < p < \alpha$ de uma v.a. $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, 0)$ com $0 < \alpha < 2$, $\alpha \neq 1$ pode ainda ser expresso por,

$$\mathbb{E}\,|\mathrm{Z}|^p = \frac{\sigma^p\,\Gamma\left(1-\frac{p}{\alpha}\right)\cos\left[\frac{p}{\alpha}\,\arctan\left(\lambda\tan\frac{\alpha\pi}{2}\right)\right]}{\Gamma(1-p)\left|\cos\left[\arctan\left(\lambda\tan\frac{\alpha\pi}{2}\right)\right]\right|^{\frac{p}{\alpha}}\cos\left(\frac{p\pi}{2}\right)}$$

(ver [7], página 38).

Proposição 1.4.8 *Se* $\mathrm{Z} \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$ *com* $1 < \alpha \leqslant 2$ *então* $\mu = \mathbb{E}(\mathrm{Z})$.

Demonstração. Consultar [82], página 19. ∎

Finalizamos esta secção com dois resultados que nos dão a equivalência para convergência quase certa de somas de v.a's independentes estáveis. O primeiro diz-nos quais as condições sobre os parâmetros que são equivalentes à convergência quase certa de uma série de v.a.'s independentes α-estáveis.

Teorema 1.4.5 *Seja* $\{\mathrm{X}_i,\ i \geqslant 1\}$ *uma sucessão de variáveis aleatórias independentes tal que* $\mathrm{X}_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, \mu_i)$, $0 < \alpha \leqslant 2$. *Então,*

$$\sum_{i=1}^\infty \sigma_i^\alpha < \infty \quad e \quad \sum_{i=1}^\infty \mu_i \ \text{converge} \qquad \Longleftrightarrow \qquad \sum_{i=1}^\infty \mathrm{X}_i \ \text{converge q.c.}$$

Demonstração. A demonstração pode ser vista em [67]. ∎

Através do teorema das três séries de Kolmogorov é ainda possível estabelecer condições para que uma série de v.a.'s independentes convirja absolutamente quase certamente.

Teorema 1.4.6 *Seja* $\{\mathrm{X}_i,\ i \geqslant 1\}$ *uma sucessão de variáveis aleatórias independentes tal que* $\mathrm{X}_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, \mu_i)$, $0 < \alpha \leqslant 2$. *Então* $\sum_{i=1}^\infty \mathrm{X}_i$ *converge absolutamente q.c. se e só se* $\sum_{i=1}^\infty |\mu_i| < \infty$ *e*

$$\begin{cases} \sum_{i=1}^\infty \sigma_i^\alpha < \infty & \text{se} \quad 0 < \sigma < 1 \\ \\ \sum_{i=1}^\infty \sigma_i \left|\log\left(\min\{\sigma_i, 0.5\}\right)\right| < \infty & \text{se} \quad \alpha = 1 \\ \\ \sum_{i=1}^\infty \sigma_i < \infty & \text{se} \quad \alpha > 1 \end{cases}.$$

Demonstração.

(a) $0 < \alpha < 1$.

Suponhamos que $\sum_{i=1}^{\infty} \sigma_i^{\alpha} < \infty$ e $\sum_{i=1}^{\infty} |\mu_i| < \infty$. Para demonstrarmos que $\sum_{i=1}^{\infty} X_i$ converge absolutamente quase certamente teremos apenas de mostrar,

$$\sum_{i=1}^{\infty} \mathbb{E}\left(|X_i|\, I_{\{|X_i| \leqslant 1\}}\right) < \infty \tag{1.4.5}$$

já que a convergência de $\sum_{i=1}^{\infty} \mathbb{P}\left(|X_i| > 1\right)$ vem directamente do Teorema 1.4.5, Teorema 1.1.19 e a convergência de $\sum_{i=1}^{\infty} \mathbb{E}\left(X_i^2 I_{\{|X_i| \leqslant 1\}}^2\right)$ segue ainda do Teorema 1.1.19 e de (1.4.5). Das propriedades das distribuições estáveis sabemos que, $X_i \sim \sigma_i Y_i + \mu_i$ com $Y_i \sim \mathcal{S}_{\alpha}(1, \lambda_i, 0)$ e

$$Y_i \sim \left(\frac{1+\lambda_i}{2}\right)^{\frac{1}{\alpha}} A_i - \left(\frac{1-\lambda_i}{2}\right)^{\frac{1}{\alpha}} B_i$$

onde A_i e B_i são independentes com distribuição comum $\mathcal{S}_{\alpha}(1, 1, 0)$. Observando que uma variável aleatória com distribuição $\mathcal{S}_{\alpha}(\sigma, 1, 0)$, $0 < \alpha < 1$ tem o suporte contido em $[0, \infty[$ (ver [82], página 15) vem

$$\begin{aligned}
\mathbb{E}\left(|X_i|\, I_{\{|X_i| \leqslant 1\}}\right) &\leqslant \int_0^1 \mathbb{P}\left(\sigma_i |Y_i| + |\mu_i| > x\right) dx \\
&= \int_0^1 \mathbb{P}\left(|Y_i| > \frac{x - |\mu_i|}{\sigma_i}\right) dx \\
&\leqslant \int_0^1 \mathbb{P}\left(|A_1| + |B_1| > \frac{x - |\mu_i|}{\sigma_i}\right) dx \\
&= \sigma_i \int_{-\frac{|\mu_i|}{\sigma_i}}^{\frac{1-|\mu_i|}{\sigma_i}} \mathbb{P}\left(A_1 + B_1 > y\right) dy \\
&= \sigma_i \int_{-\frac{|\mu_i|}{\sigma_i}}^{0} \mathbb{P}\left(A_1 + B_1 > y\right) dy + \sigma_i \int_0^{\frac{1-|\mu_i|}{\sigma_i}} \mathbb{P}\left(A_1 + B_1 > y\right) dy \\
&\leqslant |\mu_i| + \sigma_i \int_0^{\frac{1}{\sigma_i}} \mathbb{P}\left(A_1 + B_1 > y\right) dy
\end{aligned}$$

e uma vez que $\mathbb{P}\left(A_1 + B_1 > t\right) \sim 2\mathbb{P}\left(A_1 > t\right) \sim 2c(\alpha) t^{-\alpha}$ quando $t \to \infty$ (ver [26], página 37) tem-se,

$$\int_0^{\frac{1}{\sigma_i}} \mathbb{P}\left(A_1 + B_1 > y\right) dy \sim 2c(\alpha) \int_0^{\frac{1}{\sigma_i}} \frac{1}{y^{\alpha}}\, dy = \frac{2c(\alpha)}{1-\alpha} \sigma_i^{\alpha - 1}, \quad i \to \infty.$$

Para estabelecermos a implicação recíproca basta utilizar o Teorema 1.4.5 para concluir que $\sum_{i=1}^{\infty} \sigma_i^{\alpha} < \infty$. Mas se $\sum_{i=1}^{\infty} |X_i| < \infty$ quase certamente então,

$$\sum_{i=1}^{\infty} s^*(\mu_i) X_i \quad \text{converge} \quad \text{q.c.}$$

onde,

$$s^*(t) = \begin{cases} -1 & \text{se} \quad t < 0 \\ 1 & \text{se} \quad t \geqslant 0 \end{cases}$$

(uma vez que converge absolutamente quase certamente). Visto que $s^*(\mu_i)\mathrm{X}_i \sim \mathcal{S}_\alpha(\sigma_i, s^*(\mu_i)\lambda_i, |\mu_i|)$ concluímos (novamente do Teorema 1.4.5) que $\displaystyle\sum_{i=1}^{\infty} |\mu_i| < \infty$.

(b) $\alpha = 1$.

Admitamos que $\displaystyle\sum_{i=1}^{\infty} \sigma_i \left|\log\left(\min\{\sigma_i, 0.5\}\right)\right| < \infty$ e $\displaystyle\sum_{i=1}^{\infty} |\mu_i| < \infty$. À semelhança do caso anterior, teremos apenas de demonstrar (1.4.5). Das propriedades das distribuições estáveis tem-se,

$$\mathrm{X}_i \sim \sigma_i \mathrm{Y}_i + \frac{2}{\pi}\lambda_i \sigma_i \log \sigma_i + \mu_i$$

com $\mathrm{Y}_i \sim \mathcal{S}_1(1, \lambda_i, 0)$ e

$$\mathrm{Y}_i \sim \left(\frac{1+\lambda_i}{2}\right)\mathrm{A}_i - \left(\frac{1-\lambda_i}{2}\right)\mathrm{B}_i + \left(\frac{1+\lambda_i}{\pi}\right)\log\left(\frac{1+\lambda_i}{2}\right) - \left(\frac{1-\lambda_i}{\pi}\right)\log\left(\frac{1-\lambda_i}{2}\right)$$

onde A_i e B_i são independentes com distribuição comum $\mathcal{S}_1(1,1,0)$. Então,

$$\mathbb{E}\big(\, |\mathrm{X}_i|\, I_{\{|\mathrm{X}_i| \leq 1\}}\big) = \int_0^1 \mathbb{P}\left(\left|\sigma_i \mathrm{Y}_i + \frac{2}{\pi}\lambda_i\sigma_i\log\sigma_i + \mu_i\right| > x\right) dx \leqslant \int_0^1 \mathbb{P}\left(|\mathrm{Y}_i| > \frac{x - \frac{2}{\pi}\sigma_i\left|\log\sigma_i\right| - |\mu_i|}{\sigma_i}\right) dx$$

e

$$\mathbb{E}\big(\, |\mathrm{X}_i|\, I_{\{|\mathrm{X}_i| \leq 1\}}\big) \leqslant \int_0^1 \mathbb{P}\left(|\mathrm{A}_1| + |\mathrm{B}_1| > \frac{x - \frac{2}{\pi}\sigma_i\left|\log\sigma_i\right| - |\mu_i| - c\sigma_i}{\sigma_i}\right) dx \leqslant$$

$$\leqslant \sigma_i \int_{\frac{-\frac{2}{\pi}\sigma_i|\log\sigma_i| - |\mu_i| - c\sigma_i}{\sigma_i}}^{\frac{1}{\sigma_i}} \mathbb{P}\left(|\mathrm{A}_1| + |\mathrm{B}_1| > y\right) dy \qquad \left(c < \frac{1}{10}\right)$$

Visto que $\mathbb{P}\left(|\mathrm{A}_1| > t\right) = \mathbb{P}\left(\mathrm{A}_1 > t\right) + \mathbb{P}\left(\mathrm{A}_1 < -t\right) \sim \dfrac{2}{\pi}t^{-1}$, $t \to \infty$ e $\mathbb{P}\left(|\mathrm{A}_1| + |\mathrm{B}_1| > t\right) \sim 2\mathbb{P}\left(|\mathrm{A}_1| > t\right)$, $t \to \infty$ tem-se,

$$\mathbb{P}\left(|\mathrm{A}_1| + |\mathrm{B}_1| > t\right) \sim \frac{4}{\pi}t^{-1}, \quad t \to \infty.$$

Então o resultado segue pois,

$$\int_1^{\frac{1}{\sigma_i}} \mathbb{P}\left(|\mathrm{A}_1| + |\mathrm{B}_1| > y\right) dy \sim \frac{4}{\pi}\log\left(\frac{1}{\sigma_i}\right), \quad i \to \infty$$

e

$$\sigma_i \int_{\frac{-\frac{2}{\pi}\sigma_i|\log\sigma_i| - |\mu_i| - c\sigma_i}{\sigma_i}}^{\frac{1}{\sigma_i}} \mathbb{P}\left(|\mathrm{A}_1| + |\mathrm{B}_1| > y\right) dy =$$

$$= \frac{2}{\pi}\sigma_i\left|\log\sigma_i\right| + |\mu_i| + c\sigma_i + \sigma_i\int_0^1 \mathbb{P}\left(|\mathrm{A}_1| + |\mathrm{B}_1| > y\right) dy + \sigma_i \int_1^{\frac{1}{\sigma_i}} \mathbb{P}\left(|\mathrm{A}_1| + |\mathrm{B}_1| > y\right) dy.$$

Para a implicação recíproca, o argumento utilizado no caso anterior permite, desde logo, concluir que $\sum_{i=1}^{\infty} |\mu_i| < \infty$. Então, resta-nos mostrar apenas que $\sum_{i=1}^{\infty} \sigma_i \, |\log(\min\{\sigma_i, 0.5\})| < \infty$. Como $|X_i| = |X_i - \mu_i + \mu_i| \geqslant |X_i - \mu_i| - |\mu_i|$ bastará considerar $X_i \sim \mathcal{S}_1(\sigma_i, \lambda_i, 0)$ e assim

$$\begin{aligned} X_i \sim \sigma_i \left(\frac{1+\lambda_i}{2}\right) A_i - \sigma_i \left(\frac{1-\lambda_i}{2}\right) B_i + \\ + \sigma_i \underbrace{\left(\frac{1+\lambda_i}{\pi}\right) \log\left(\frac{1+\lambda_i}{2}\right) - \left(\frac{1-\lambda_i}{\pi}\right) \log\left(\frac{1-\lambda_i}{2}\right)}_{=\zeta_i} + \frac{2}{\pi} \lambda_i \sigma_i \log \sigma_i \end{aligned}$$

onde A_i e B_i são independentes com distribuição comum $\mathcal{S}_1(1,1,0)$. Deste modo,

$$\mathbb{E}\left(|X_i| \, I_{\{|X_i| \leqslant 1\}}\right) \geqslant \sigma_i \cdot \frac{\mathbb{P}(A_1 < 0)}{2} \cdot \left(\int_1^{\frac{1-\frac{\sigma_i}{10}}{\sigma_i}} \mathbb{P}(A_1 > z) \, dz - \int_1^{-\frac{5}{\pi}\log \sigma_i} \mathbb{P}(A_1 > z) \, dz \right)$$

pois se $0 < \lambda_i < 1$ (os restantes casos são análogos) tem-se,

$$\begin{aligned} \mathbb{E}\left(|X_i| \, I_{\{|X_i| \leqslant 1\}}\right) &= \int_0^1 \mathbb{P}(|X_i| > x) \, dx \\ &\geqslant \int_0^1 \mathbb{P}\left(\sigma_i \left(\frac{1+\lambda_i}{2}\right) A_i - \sigma_i \left(\frac{1-\lambda_i}{2}\right) B_i > x - \sigma_i \zeta_i - \frac{2}{\pi}\lambda_i \sigma_i \log \sigma_i \right) dx \\ &= \int_{-\sigma_i \zeta_i - \frac{2}{\pi}\lambda_i\sigma_i \log\sigma_i}^{1-\sigma_i \zeta_i - \frac{2}{\pi}\lambda_i\sigma_i \log\sigma_i} \mathbb{P}\left(\sigma_i \left(\frac{1+\lambda_i}{2}\right) A_i - \sigma_i \left(\frac{1-\lambda_i}{2}\right) B_i > y\right) dy \\ &\underset{\substack{\mathbb{P}(U+V>x) \geqslant \mathbb{P}(U>x)\cdot\mathbb{P}(V>0) \\ \text{com U e V v.a.'s independentes}}}{\geqslant} \int_{-\sigma_i \zeta_i - \frac{2}{\pi}\lambda_i\sigma_i \log\sigma_i}^{1-\sigma_i \zeta_i - \frac{2}{\pi}\lambda_i\sigma_i \log\sigma_i} \mathbb{P}\left(\sigma_i \left(\frac{1+\lambda_i}{2}\right) A_i > y\right) \cdot \mathbb{P}\left(-\sigma_i \left(\frac{1-\lambda_i}{2}\right) B_i > 0\right) dy \\ &= \sigma_i \cdot \left(\frac{1+\lambda_i}{2}\right) \cdot \mathbb{P}(B_1 < 0) \cdot \int_{\frac{-\sigma_i \zeta_i - \frac{2}{\pi}\lambda_i\sigma_i \log\sigma_i}{\sigma_i\left(\frac{1+\lambda_i}{2}\right)}}^{\frac{1-\sigma_i \zeta_i - \frac{2}{\pi}\lambda_i\sigma_i \log\sigma_i}{\sigma_i\left(\frac{1+\lambda_i}{2}\right)}} \mathbb{P}(A_1 > z) \, dz \\ &\underset{\substack{\frac{1+\lambda_i}{2} \geqslant 0, \ \forall i \\ -\frac{1}{10} \leqslant \zeta_i \leqslant \frac{1}{10}, \ \forall i \\ -\frac{2}{\pi}\lambda_i\sigma_i\log\sigma_i > 0, \ \forall i \geqslant i_1}}{\geqslant} \sigma_i \cdot \frac{\mathbb{P}(B_1 < 0)}{2} \cdot \int_{\frac{1}{5} - \frac{4}{\pi}\log\sigma_i}^{\frac{1-\frac{\sigma_i}{10}}{\sigma_i}} \mathbb{P}(A_1 > z) \, dz \\ &\underset{-\frac{1}{\pi}\log\sigma_i \geqslant \frac{1}{5}, \ \forall i \geqslant i_2}{\geqslant} \sigma_i \cdot \frac{\mathbb{P}(A_1 < 0)}{2} \cdot \int_{-\frac{5}{\pi}\log\sigma_i}^{\frac{1-\frac{\sigma_i}{10}}{\sigma_i}} \mathbb{P}(A_1 > z) \, dz \\ &= \sigma_i \cdot \frac{\mathbb{P}(A_1 < 0)}{2} \cdot \left(\int_1^{\frac{1-\frac{\sigma_i}{10}}{\sigma_i}} \mathbb{P}(A_1 > z) \, dz - \int_1^{-\frac{5}{\pi}\log\sigma_i} \mathbb{P}(A_1 > z) \, dz \right). \end{aligned}$$

Tendo em conta que $\mathbb{P}(A_1 < 0)$ é não-nulo (pois o suporte da distribuição $\mathcal{S}_1(1,1,0)$ é $\mathbb{R}$) o resultado fica estabelecido já que,

$$\int_1^{\frac{1-\frac{\sigma_i}{10}}{\sigma_i}} \mathbb{P}(A_1 > z) \, dz \sim \log\left(\frac{1}{\sigma_i}\right), \quad i \to \infty \quad \text{e} \quad \int_1^{-\frac{5}{\pi}\log\sigma_i} \mathbb{P}(A_1 > z) \, dz = o\big(\log\sigma_i\big), \quad i \to \infty.$$

(c) $1 < \alpha < 2$.

Suponhamos que $\sum_{i=1}^{\infty} \sigma_i < \infty$ e $\sum_{i=1}^{\infty} |\mu_i| < \infty$. Então,

$$\mathbb{E}\,|\mathrm{X}_i| = \mathbb{E}\,|\mathrm{X}_i - \mathbb{E}(\mathrm{X}_i) + \mu_i| \leqslant \mathbb{E}\,|\mathrm{X}_i - \mathbb{E}(\mathrm{X}_i)| + |\mu_i| = c(\alpha, \lambda_i)\sigma_i + |\mu_i|$$

com $c(\alpha, \lambda_i) \leqslant \frac{\Gamma\left(1-\frac{1}{\alpha}\right)}{\int_0^\infty u^{-2}\sin^2 u\, du}\left[1+\tan^2\left(\frac{\alpha\pi}{2}\right)\right]^{\frac{1}{2\alpha}}$ donde $\sum_{i=1}^{\infty} \mathbb{E}\,|\mathrm{X}_i| < \infty$. Consequentemente,

$$\sum_{i=1}^{\infty} |\mathrm{X}_i| < \infty \quad \text{q.c.}$$

(ver [57], página 80).

Reciprocamente, se $\sum_{i=1}^{\infty} |\mathrm{X}_i| < \infty$ quase certamente então o mesmo argumento que foi utilizado nos casos anteriores permite concluir que $\sum_{i=1}^{\infty} |\mu_i| < \infty$. Para vermos que $\sum_{i=1}^{\infty} \sigma_i < \infty$ consideremos novamente $\mathrm{X}_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, 0)$. Então,

$$\mathrm{X}_i \sim \sigma_i\left(\frac{1+\lambda_i}{2}\right)^{\frac{1}{\alpha}}\mathrm{A}_i - \sigma_i\left(\frac{1-\lambda_i}{2}\right)^{\frac{1}{\alpha}}\mathrm{B}_i$$

onde A_i e B_i são independentes com distribuição comum $\mathcal{S}_\alpha(1,1,0)$. Deste modo,

$$\mathbb{E}\left(|\mathrm{X}_i|\, I_{\{|\mathrm{X}_i| \leqslant 1\}}\right) \geqslant \sigma_i \cdot \frac{\mathbb{P}\left(\mathrm{A}_1 < 0\right) \cdot \mathbb{P}\left(\mathrm{A}_1 > 1\right)}{2^{\frac{1}{\alpha}}}$$

porque se $-1 < \lambda_i < 0$ (os restantes casos são análogos) tem-se,

$$\begin{aligned}
\mathbb{E}\left(|\mathrm{X}_i|\, I_{\{|\mathrm{X}_i| \leqslant 1\}}\right) &= \int_0^1 \mathbb{P}\left(|\mathrm{X}_i| > x\right)\, dx \\
&\geqslant \int_0^1 \mathbb{P}\left(\sigma_i\left(\frac{1+\lambda_i}{2}\right)^{\frac{1}{\alpha}}\mathrm{A}_i - \sigma_i\left(\frac{1-\lambda_i}{2}\right)^{\frac{1}{\alpha}}\mathrm{B}_i < -x\right) dx \\
&= \int_0^1 \mathbb{P}\left(-\sigma_i\left(\frac{1+\lambda_i}{2}\right)^{\frac{1}{\alpha}}\mathrm{A}_i + \sigma_i\left(\frac{1-\lambda_i}{2}\right)^{\frac{1}{\alpha}}\mathrm{B}_i > x\right) dx \\
&\geqslant \int_0^1 \mathbb{P}\left(-\sigma_i\left(\frac{1+\lambda_i}{2}\right)^{\frac{1}{\alpha}}\mathrm{A}_i > 0\right) \cdot \mathbb{P}\left(\sigma_i\left(\frac{1-\lambda_i}{2}\right)^{\frac{1}{\alpha}}\mathrm{B}_i > x\right) dx \\
&= \sigma_i \cdot \left(\frac{1-\lambda_i}{2}\right)^{\frac{1}{\alpha}} \cdot \mathbb{P}\left(\mathrm{A}_1 < 0\right) \cdot \int_0^{\frac{1}{\sigma_i\left(\frac{1-\lambda_i}{2}\right)^{\frac{1}{\alpha}}}} \mathbb{P}\left(\mathrm{B}_1 > y\right)\, dy \\
&\geqslant \sigma_i \cdot \frac{\mathbb{P}\left(\mathrm{A}_1 < 0\right)}{2^{\frac{1}{\alpha}}} \cdot \int_0^{\frac{1}{\sigma_i}} \mathbb{P}\left(\mathrm{B}_1 > y\right)\, dy \\
&= \sigma_i \cdot \frac{\mathbb{P}\left(\mathrm{A}_1 < 0\right)}{2^{\frac{1}{\alpha}}} \cdot \left(\int_0^1 \mathbb{P}\left(\mathrm{B}_1 > y\right)\, dy + \int_1^{\frac{1}{\sigma_i}} \mathbb{P}\left(\mathrm{B}_1 > y\right)\, dy\right) \\
&\geqslant \sigma_i \cdot \frac{\mathbb{P}\left(\mathrm{A}_1 < 0\right)}{2^{\frac{1}{\alpha}}} \cdot \int_0^1 \mathbb{P}\left(\mathrm{B}_1 > y\right)\, dy
\end{aligned}$$

$$\underset{y\mapsto\mathbb{P}(\mathrm{B}_1>y)\ \text{decrescente}}{\geqslant} \sigma_i\cdot\frac{\mathbb{P}\left(\mathrm{A}_1<0\right)\cdot\mathbb{P}\left(\mathrm{B}_1>1\right)}{2^{\frac{1}{\alpha}}}$$
$$=\sigma_i\cdot\frac{\mathbb{P}\left(\mathrm{A}_1<0\right)\cdot\mathbb{P}\left(\mathrm{A}_1>1\right)}{2^{\frac{1}{\alpha}}}.$$

O resultado fica estabelecido uma vez que $\mathbb{P}\left(\mathrm{A}_1<0\right)$ ou $\mathbb{P}\left(\mathrm{A}_1>1\right)$ são não-nulos (o suporte da distribuição $\mathcal{S}_\alpha(1,1,0)$, $\alpha>1$ é $\mathbb{R}$).

(d) $\alpha=2$.

Se $\mathrm{X}_i\sim N(\mu_i,2\sigma_i^2)$ então $\mathrm{X}_i=\sqrt{2}\sigma_i\mathrm{Z}_i+\mu_i$ com $\mathrm{Z}_i\sim N(0,1)$, $\forall i$ e a demonstração segue os mesmos passos do caso acima. ■

1.5 Domínios de atracção maximal

No mundo que nos rodeia, existe um vasto número de situações onde os extremos governam as leis que nos interessam conhecer. Alguns casos concretos (ver [31]) são listados de seguida:

Desastres naturais. Inundações, chuvas torrenciais, temperaturas extremas, pressões atmosféricas extremas, ventos ciclónicos e outros fenómenos podem causar extensivas perdas humanas e materiais. Enquanto tais desastres não puderem ser completamente evitados, a sociedade deve tomar acções de prevenção no sentido de minimizar os estragos. Em represas, diques, canais e outras estruturas a escolha dos materiais de construção e dos métodos de arquitectura podem acautelar alguns destes desastres. As soluções, dadas pela engenharia, que tentam responder a estes problemas devem basear-se numa teoria bastante precisa já que falhas de imprecisão podem revelar-se fatais (por exemplo, represas de grande custo podem não durar o suficiente antes de entrarem em colapso).

Falha numa peça de um equipamento. Suponhamos que a peça de um equipamento falha se uma das suas componentes falhar. Por outras palavras, consideremos o conjunto de componentes dessa peça, cuja a falha de qualquer uma delas conduz à paragem do mecanismo. Esta é uma situação extrema, no sentido em que a "componente mais fraca"por si só leva à falha do equipamento. Apesar desta suposição parecer simplista, o modelo geral para a falha duma complexa peça de um equipamento pode ser reduzido ao modelo que acabamos de descrever. Na realidade, se primeiro considerarmos grupos de componentes onde a falha de um deles resulta na paragem do mecanismo então o "grupo mais fraco"terá influência na primeira falha no equipamento.

Tempo de serviço. Consideremos, novamente, uma peça de um equipamento com um largo número de componentes e admitamos que essas componentes servem o sistema em simultâneo. Então, o tempo necessário para que o equipamento seja servido é determinado pela componente que requer o tempo de serviço mais longo.

Corrosão. Costumamos dizer que uma superfície com um largo número de pequenas cavidades entra em corrosão se qualquer uma delas penetrar através da espessura da superfície. Inicialmente, as cavidades são de profundidade aleatória, a qual, devido à corrosão química, vai crescendo com o tempo. Mais uma vez, estamos perante uma medição extrema que causa a falha: a profundidade do buraco.

Solidez de um material. Um material absolutamente homogéneo quebra sob pressão obedecendo a uma certa lei determinística. Contudo, nenhum material é absolutamente homogéneo; de facto, experiências de engenharia mostram que a quebra de materiais sólidos produzidos sob idênticos procedimentos varia grandemente. A explicação é que cada ponto (ou pelo menos cada pequena área) tem uma solidez aleatória e portanto, a quantidade de força que é necessária para a quebra do material varia nos diferentes pontos. Evidentemente, o "ponto mais fraco"determinará a solidez de todo o material.

Poluição atmosférica. A concentração de ar poluído é expressa em termos da proporção de um poluente específico existente no ar. As concentrações são registadas em intervalos de tempo iguais (as investigações actuais baseiam-se em informações recolhidas em intervalos de cinco minutos) e o objectivo é manter a "maior medição"abaixo dos níveis aceitáveis.

Amostras estatísticas. Efectuadas as observações de uma determinada quantidade queremos, muitas vezes, saber quão grande ou quão pequena uma medição pode ser.

Estimadores estatísticos. Depois de se recolherem as observações, as informações são usadas para calcular estimadores de certas características de uma determinada quantidade sob observação. Consequentemente, pretendemos estimar essas características de uma forma tão precisa quanto possível mas estimações acima ou abaixo serão inevitáveis. De considerável interesse passa a ser, então, a investigação daquele que é o maior ou menor estimador.

Os problemas associados aos extremos, que não se esgotam nos exemplos acima descritos, indicam que qualquer teoria de extremos bem sucedida reunirá um extenso número de tópicos interessantes das mais diversas áreas do conhecimento. Matematicamente, todos os problemas que foram referidos, podem ser interpretados formalmente do seguinte modo: efectuado um número n de medições aleatórias $\mathrm{X}_1, \ldots, \mathrm{X}_n$ pretende-se estudar e conhecer o comportamento de $\min(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ ou de $\max(\mathrm{X}_1, \ldots, \mathrm{X}_n)$. Se pensarmos numa inundação provocada pelo excesso de caudal de um rio, X_j indicará o nível do rio no dia j onde, por exemplo, podemos considerar como "dia 1"o dia de hoje. Visto que desconhecemos o nível do rio no futuro, podemos tomá-lo como aleatório. A resposta à questão "Estará o nível do rio abaixo de 250cm daqui a 4 anos?"passa pelo cálculo de $\mathbb{P}\big\{\max(\mathrm{X}_1, \ldots, \mathrm{X}_n) \leqslant 250\big\}$ em que $n = 1461$.

Mais concretamente, o objectivo será impor condições sobre sucessões de constantes $b_n, d_n \in \mathbb{R}$ e $a_n, c_n > 0$ tais que,

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{\min(\mathrm{X}_1, \ldots, \mathrm{X}_n) - b_n}{a_n} \leqslant x\right\} = L(x)$$

e

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{\max\left(\mathrm{X}_1,\ldots,\mathrm{X}_n\right)-d_n}{c_n}\leqslant x\right\}=H(x)$$

para algumas funções de distribuição $L(x)$ e $H(x)$ a determinar. Embora hajam resultados de convergência desenvolvidos num cenário mais geral no que concerne às hipóteses sobre as v.a.'s $\mathrm{X}_1,\ldots,\mathrm{X}_n$ (ver [31]) assumiremos sempre que estas v.a.'s são i.i.d. não-degeneradas. Convém ainda sublinhar a importância do estudo assimptótico do modelo já que, em muitas situações, a análise assimptótica conduz-nos ao modelo exacto para o fenómeno em estudo, enquanto um número fixo de observações apenas pode utilizado como uma aproximação. A espectacularidade desta teoria reside no facto de virmos a compreender a regularidade de comportamento de fenómenos extremos... o que até parece contraditório!

Ao longo desta secção $\mathrm{X},\mathrm{X}_1,\mathrm{X}_2,\ldots$ será uma sucessão de v.a.'s i.i.d. não-degeneradas com função de distribuição comum F. Por uma questão de necessidade de resultados específicos em capítulos futuros, trabalharemos doravante, apenas e só com $\max\left(\mathrm{X}_1,\ldots,\mathrm{X}_n\right)$. No entanto, todos os resultados desenvolvidos para $\max\left(\mathrm{X}_1,\ldots,\mathrm{X}_n\right)$ possuem versões análogas para $\min\left(\mathrm{X}_1,\ldots,\mathrm{X}_n\right)$ dada a relação $\min\left(\mathrm{X}_1,\ldots,\mathrm{X}_n\right)=-\max\left(-\mathrm{X}_1,\ldots,-\mathrm{X}_n\right)$. Denotemos,

$$\mathrm{M}_1=\mathrm{X}_1,\quad \mathrm{M}_n=\max\left(\mathrm{X}_1,\ldots,\mathrm{X}_n\right)\qquad (n\geqslant 2)$$

cuja função de distribuição é $\mathbb{P}\{\mathrm{M}_n\leqslant x\}=\mathbb{P}\{\mathrm{X}_1\leqslant x,\ldots,\mathrm{X}_n\leqslant x\}=F^n(x)$, $x\in\mathbb{R}$, $n\in\mathbb{N}$. O tópico principal desta secção consistirá em caracterizar todas as possíveis distribuições limite da sucessão centralizada e normalizada $c_n^{-1}\big(\mathrm{M}_n-d_n\big)$, $c_n>0$, $d_n\in\mathbb{R}$. Esta questão está intimamente ligada a uma outra: quais as distribuições que satisfazem para todo o $n\geqslant 2$ a identidade em lei,

$$\max\left(\mathrm{X}_1,\ldots,\mathrm{X}_n\right)\overset{\mathrm{d}}{=}c_n\mathrm{X}+d_n$$

para sucessões de constantes apropriadas $c_n>0$ e $d_n\in\mathbb{R}$?

Definição 1.5.1 Uma v.a. não-degenerada X (ou a correspondente f.d.) diz-se **max-estável** se satisfaz,

$$\max\left(\mathrm{X}_1,\ldots,\mathrm{X}_n\right)\overset{\mathrm{d}}{=}c_n\mathrm{X}+d_n,\quad \forall n\geqslant 2 \tag{1.5.1}$$

para $\mathrm{X}_1,\ldots,\mathrm{X}_n$ cópias independentes de X e sucessões de constantes $c_n>0$, $d_n\in\mathbb{R}$ apropriadas.

As sucessões de constantes $c_n>0$ e $d_n\in\mathbb{R}$ são designadas, respectivamente, por **constantes normalizantes** e **constantes centralizantes**. Se $\mathrm{X},\mathrm{X}_1,\mathrm{X}_2,\ldots$ for uma sucessão de v.a.'s i.i.d. max-estáveis então (1.5.1) pode ser reescrita como,

$$c_n^{-1}\big(\mathrm{M}_n-d_n\big)\overset{\mathrm{d}}{=}\mathrm{X}$$

donde qualquer f.d. max-estável é distribuição limite do (apropriadamente centralizado e normalizado) máximo de v.a.'s i.i.d.; além disso, funções de distribuição max-estáveis são as únicas distribuições limite de $c_n^{-1}\big(\mathrm{M}_n - d_n\big)$. À semelhança do que já aconteceu anteriormente omitiremos as demonstrações de todos os resultados apresentados; contudo, o leitor poderá encontrá-las em livros da especialidade como [26], [31], [77] ou [78].

Teorema 1.5.1 *A classe de funções de distribuição max-estáveis coincide com a classe de todas as possíveis funções de distribuição não-degeneradas que são limite da (apropriadamente centralizada e normalizada) sucessão do máximo de v.a.'s i.i.d..*

Demonstração. Consultar [26], página 121. ∎

Definição 1.5.2 Sendo $\alpha > 0$ chamam-se **distribuições de valores extremos** às funções de distribuição Φ_α, Λ e Ψ_α definidas como se segue:

$$\Phi_\alpha(x) = \begin{cases} \exp(-x^{-\alpha}) & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

$$\Lambda(x) = \exp\big(-e^{-x}\big), \quad x \in \mathbb{R}$$

$$\Psi_\alpha(x) = \begin{cases} 1 & \text{se} \quad x > 0 \\ \\ \exp\big(-(-x)^{\alpha}\big) & \text{se} \quad x \leqslant 0 \end{cases}$$

chamadas, respectivamente, **distribuição de Fréchet**, **distribuição de Gumbel** e **distribuição de Weibull**.

O próximo resultado, o teorema de Fisher-Tippett, especifica a forma da distribuição limite para o máximo de v.a.'s i.i.d. centralizadas e normalizadas constituindo, assim, um dos resultados fundamentais da teoria de valores extremos.

Teorema 1.5.2 (teorema de Fisher-Tippett) *Seja $\{\mathrm{X}_n,\, n \geqslant 1\}$ uma sucessão de v.a.'s i.i.d.. Se existem constantes normalizantes $c_n > 0$, constantes centralizantes $d_n \in \mathbb{R}$ e alguma função de distribuição H não-degenerada tais que,*

$$\lim_{n\to\infty} F^n(c_n x + d_n) = H(x)$$

então H é do mesmo tipo de uma das três distribuições de valores extremos.

Demonstração. Consultar [26], página 121. ∎

Observação 1.5.1 Observemos que as funções de distribuição que são do mesmo tipo de uma das três distribuições de valores extremos são contínuas em todo o $\mathbb{R}$ donde podemos escrever mais correctamente: $\lim\limits_{n\to\infty} F^n(c_n x + d_n) = H(x)$, $x \in \mathbb{R}$.

Identificada que está a f.d. limite da sucessão centralizada e normalizada do máximo de v.a.'s i.i.d. levantam-se agora outras questões: dada uma distribuição de valores extremos H, que condições sobre F implicam a convergência em lei de $c_n^{-1}\big(\mathrm{M}_n - d_n\big)$? Como escolher as constantes normalizantes $c_n > 0$ e centralizantes $d_n \in \mathbb{R}$ de modo a que $\lim\limits_{n\to\infty} F^n(c_n x + d_n) = H(x)$? Pode acontecer que diferentes constantes normalizantes e centralizantes impliquem a convergência para distintas funções de distribuição limite? Esta última questão pode ser respondida de imediato: a distribuição limite é unicamente determinada a menos de uma transformação afim. Para responder às duas primeiras questões é necessário introduzir a seguinte definição.

Definição 1.5.3 Dizemos que a variável aleatória X (ou a sua função de distribuição F) pertence ao **domínio de atracção maximal** de uma distribuição H não-degenerada e escrevemos $\mathrm{X} \in \mathcal{D}(H)$ (ou $F \in \mathcal{D}(H)$) se existirem constantes normalizantes $c_n > 0$ e constantes centralizantes $d_n \in \mathbb{R}$ tais que,

$$\lim_{n\to\infty} F^n(c_n x + d_n) = H(x).$$

Até ao final desta secção, iremos dar as condições necessárias e suficientes para que a função de distribuição F pertença ao domínio de atracção maximal de uma das três distribuições de valores extremos, caracterizando em simultâneo as constantes normalizantes $c_n > 0$ e constantes centralizantes $d_n \in \mathbb{R}$. Antes, porém, será necessário introduzirmos o conceito de *função quantil* de uma distribuição.

Definição 1.5.4 Sendo F uma função de distribuição então a função dada por,

$$F^{\leftarrow}(t) = \inf\big\{x \in \mathbb{R}:\ F(x) \geqslant t\big\}, \quad 0 < t < 1$$

designa-se por **função quantil** da função de distribuição F. A quantidade $x_t = F^{\leftarrow}(t)$ chama-se ***t*-quantil** de F.

A. O domínio de atracção maximal da distribuição de Fréchet

O próximo resultado caracteriza o domínio de atracção maximal da distribuição de Fréchet Φ_α, $\alpha > 0$.

Teorema 1.5.3 (Domínio de atracção maximal de $\boldsymbol{\Phi_\alpha}$) *A função de distribuição F pertence ao domínio de atracção maximal de Φ_α, $\alpha > 0$ se e só se $\overline{F}(x) = x^{-\alpha}\ell(x)$ para alguma função de variação lenta ℓ. Se F pertence ao domínio de atracção maximal de Φ_α, $\alpha > 0$ então,*

$$\lim_{n\to\infty} F^n(c_n x) = \Phi_\alpha(x), \quad x \in \mathbb{R}$$

onde a sucessão de constantes normalizantes é $c_n = F^{\leftarrow}(1 - n^{-1})$.

Demonstração. Consultar [26], página 131. ∎

Sinteticamente, podemos escrever,

$$F \in \mathcal{D}(\Phi_\alpha) \quad \Longleftrightarrow \quad \overline{F} \in \mathcal{R}_{-\alpha}.$$

Observação 1.5.2 Observe-se que qualquer função de distribuição F que pertença ao domínio de atracção maximal de Φ_α, $\alpha > 0$ tem o extremo direito do suporte infinito, ou seja, $x_{\sup} = \infty$. Além disso, a sucessão de constantes normalizantes pode ser expressa por $c_n = n^{1/\alpha}\ell_0(n)$ para alguma função de variação lenta ℓ_0.

Num sentido eminentemente prático, tendo como objectivo as aplicações, exibimos condições suficientes para a identificação das distribuições que estão na classe acima referida.

Corolário 1.5.1 *Seja F uma função de distribuição absolutamente contínua com densidade f positiva nalgum intervalo $]x_0, \infty[$, $x_0 \in \mathbb{R}$.*

(a) *Se para algum $\alpha > 0$,*

$$\lim_{x\to\infty} \frac{x\, f(x)}{\overline{F}(x)} = \alpha$$

então $F \in \mathcal{D}(\Phi_\alpha)$ e podemos escolher a sucessão normalizante c_n de modo que $c_n f(c_n) \sim \alpha/n$, $n \to \infty$.

(b) *Se f é não-crescente e $F \in \mathcal{D}(\Phi_\alpha)$ então $\lim\limits_{x\to\infty} \frac{x\, f(x)}{\overline{F}(x)} = \alpha > 0$.*

(c) $\lim\limits_{x\to\infty} \frac{x\, f(x)}{\overline{F}(x)} = \alpha > 0$ *se e só se para algum x_0,*

$$\overline{F}(x) = c \exp\left(-\int_{x_0}^{x} \frac{\alpha(t)}{t}\, dt\right), \quad x > x_0 \qquad (c > 0)$$

onde $\alpha(x)$ é uma função mensurável definida em $]x_0, \infty[$ satisfazendo $\lim\limits_{t\to\infty} \alpha(t) = \alpha$.

Demonstração. Consultar [78], página 63. ∎

Proposição 1.5.1 *Se X é uma v.a. com função de distribuição $F \in \mathcal{D}(\Phi_\alpha)$ então $\mathbb{E}\big[(\mathrm{X}^+)^p\big] < \infty$ para qualquer $0 < p < \alpha$ onde $\mathrm{X}^+ = \max(\mathrm{X}, 0)$.*

Demonstração. Consultar [31], página 120. ∎

B. O domínio de atracção maximal da distribuição de Weibull

As funções de distribuição Ψ_α e Φ_α estão intimamente relacionadas uma vez que $\Psi_\alpha(-x^{-1}) = \Phi_\alpha(x)$, $x > 0$. Deste modo, também os respectivos domínios de atracção maximal vão estar intimamente relacionados como mostra o seguinte resultado.

Teorema 1.5.4 (Domínio de atracção maximal de Ψ_α) *A função de distribuição F com extremo direito do suporte $x_{\sup}$ pertence ao domínio de atracção maximal de Ψ_α, $\alpha > 0$ se e só se $x_{\sup} < \infty$ e $\overline{F}(x_{\sup} - x^{-1}) = x^{-\alpha}\ell(x)$ para alguma função de variação lenta ℓ. Se F pertence ao domínio de atracção maximal de Ψ_α, $\alpha > 0$ então,*

$$\lim_{n\to\infty} F^n(x_{\sup} + c_n x) = \Psi_\alpha(x), \quad x \in \mathbb{R}$$

onde a sucessão de constantes normalizantes é $c_n = x_{\sup} - F^{\leftarrow}(1 - n^{-1})$ e a sucessão de constantes centralizantes é $d_n = x_{\sup}$.

Demonstração. Consultar [26], página 135. ∎

Simbolicamente, podemos escrever,

$$F \in \mathcal{D}(\Psi_\alpha) \qquad \Longleftrightarrow \qquad x_{\sup} < \infty \quad \text{e} \quad \overline{F}(x_{\sup} - x^{-1}) \in \mathcal{R}_{-\alpha}.$$

O Corolário 1.5.1 admite um resultado correspondente para o domínio de atracção maximal da distribuição de Weibull.

Corolário 1.5.2 *Seja F uma função de distribuição absolutamente contínua com extremo direito do suporte $x_{\sup} < \infty$ e densidade f positiva nalgum intervalo $]x_0, x_{\sup}[$, $x_0 < x_{\sup}$.*

(a) *Se para algum $\alpha > 0$,*

$$\lim_{x\to x_{\sup}} \frac{(x_{\sup} - x)\, f(x)}{\overline{F}(x)} = \alpha$$

então $F \in \mathcal{D}(\Psi_\alpha)$.

(b) *Se f é não-crescente e $F \in \mathcal{D}(\Psi_\alpha)$ então $\lim_{x\to x_{\sup}} \frac{(x_{\sup}-x)\, f(x)}{\overline{F}(x)} = \alpha > 0$.*

(c) $\lim_{x\to x_{\sup}} \frac{(x_{\sup}-x)\, f(x)}{\overline{F}(x)} = \alpha > 0$ *se e só se*

$$\overline{F}(x) = c\exp\left(-\int_{x_{\sup}^{-1}}^{x} \frac{\delta(t)}{x_{\sup} - t}\, dt\right), \quad x_0 < x < x_{\sup} \qquad (c > 0)$$

onde a função $\delta\colon]0,\infty[\to]0,\infty[$ verifica $\lim_{t\to x_{\sup}} \delta(t) = \alpha$.

Demonstração. Consultar [78], página 63. ∎

C. O domínio de atracção maximal da distribuição de Gumbel

Finalizamos esta secção com a análise da terceira grande classe de distribuições: o domínio de atracção maximal da distribuição de Gumbel. O domínio de atracção maximal da distribuição de Gumbel cobre um vasto conjunto de distribuições, i.e. distribuições com comportamentos de cauda variadamente distintos. De entre estas, encontram-se distribuições cuja cauda de probabilidade decresce para zero mais rapidamente que qualquer potência $x^{-\alpha}$, em especial, a distribuição log-normal (dita de cauda *moderadamente pesada*) e a distribuição normal (chamada de cauda *light*). No entanto, convém destacar que existem distribuições com suporte limitado à direita que estão também incluídas nesta classe.

Definiremos a **função auxiliar** de uma função de distribuição F por,

$$A(x) = \frac{1}{\overline{F}(x)} \int_x^{x_{\text{sup}}} \overline{F}(t)\, dt, \quad x < x_{\text{sup}}.$$

Refira-se que, dada uma variável aleatória X com função de distribuição F a função auxiliar $A(x)$ não é mais do que a **função excesso média**: $A(x) = \mathbb{E}\big(\mathrm{X} - x | \mathrm{X} > x\big)$, $x < x_{\text{sup}}$ também conhecida por **função vida residual esperada** quando $\mathrm{X} \geqslant 0$ e $\mathbb{E}(\mathrm{X}) < \infty$.

Os dois próximos resultados caracterizam por completo as funções de distribuição desta vasta classe.

Teorema 1.5.5 (Domínio de atracção maximal de Λ) *A função de distribuição F com extremo direito do suporte $x_{\text{sup}} \leqslant \infty$ pertence ao domínio de atracção maximal de Λ se e só se existir algum $x_0 < x_{\text{sup}}$ tal que,*

$$\overline{F}(x) = h(x) \exp\left(-\int_{x_0}^{x} \frac{g(t)}{a(t)}\, dt\right), \quad x_0 < x < x_{\text{sup}}$$

onde $h(x)$, $g(x)$ são funções mensuráveis que satisfazem $\lim_{x \to x_{\text{sup}}} h(x) = h_0 > 0$, $\lim_{x \to x_{\text{sup}}} g(x) = 1$ e $a(x)$ é uma função positiva, absolutamente contínua em $]x_0, x_{\text{sup}}[$[22] *com derivada $a'(x)$ verificando $\lim_{x \to x_{\text{sup}}} a'(x) = 0$. A sucessão de constantes centralizantes pode ser escolhida como $d_n = F^{\leftarrow}(1 - n^{-1})$ e a sucessão de constantes normalizantes por $c_n = a(d_n)$.*

Demonstração. Consultar [26], página 142. ∎

A outra caracterização é expressa no,

Teorema 1.5.6 *A função de distribuição F com extremo direito do suporte $x_{\text{sup}} \leqslant \infty$ pertence ao domínio de atracção maximal de Λ se e só se existe alguma função positiva $\hat{a}$ tal que,*

$$\lim_{x \to x_{\text{sup}}} \frac{\overline{F}\big(x + t\hat{a}(x)\big)}{\overline{F}(x)} = e^{-t}, \quad t \in \mathbb{R}.$$

Demonstração. Consultar [26], página 143. ∎

Observação 1.5.3 Observe-se que nos teoremas anteriores, a função auxiliar $A(x)$ é uma escolha possível para ambas as funções $a(x)$ e $\hat{a}(x)$.

[22] Seja f uma função real de variável real definida num intervalo J de $\mathbb{R}$. Se para todo o $\delta > 0$ existir um $\varepsilon > 0$ tal que $\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \delta$ para cada família finita $\big\{]a_i, b_i[,\ i \geqslant 1\big\}$ de subintervalos abertos de J disjuntos dois a dois que satisfaça $\sum_{i=1}^{n} (b_i - a_i) < \varepsilon$ dizemos que f é **absolutamente contínua** em J (ver [39], página 282).

Em particular, no domínio de atracção maximal da distribuição de Gumbel encontram-se as **funções de von Mises**, isto é, as funções F que admitem a seguinte representação,

$$\overline{F}(x) = c \exp\left\{-\int_{x_0}^{x} \frac{1}{a(t)}\, dt\right\}, \quad x_0 < x < x_{\sup} \quad (c > 0)$$

com $a(x)$ uma função positiva, absolutamente contínua em $]x_0, x_{\sup}[$ com derivada $a'(x)$ verificando $\lim\limits_{x \to x_{\sup}} a'(x) = 0$.

As versões dos anteriores corolários dos domínios de atracção maximal da distribuição de Fréchet e Weibull possuem, no domínio de atracção maximal da distribuição de Gumbel, dois resultados correspondentes consoante a existência ou não de segunda derivada da função de distribuição F.

Corolário 1.5.3 *Seja F uma função de distribuição absolutamente contínua com extremo direito do suporte $x_{\sup} \leqslant \infty$ e densidade f.*

(a) *Se*

$$\lim_{x \to x_{\sup}} \frac{f(x)}{\left[\overline{F}(x)\right]^2} \int_x^{x_{\sup}} \overline{F}(t)\, dt = 1$$

então $F \in \mathcal{D}(\Lambda)$. A sucessão de constantes centralizantes pode ser escolhida como $d_n = F^{\leftarrow}(1 - n^{-1})$ e a sucessão de constantes normalizantes por $c_n = A(d_n)$.

(b) *Se f é não-crescente e $F \in \mathcal{D}(\Lambda)$ então $\lim\limits_{x \to x_{\sup}} \frac{f(x)}{\left[\overline{F}(x)\right]^2} \int_x^{x_{\sup}} \overline{F}(t)\, dt = 1$.*

(c) $\lim\limits_{x \to x_{\sup}} \frac{f(x)}{\left[\overline{F}(x)\right]^2} \int_x^{x_{\sup}} \overline{F}(t)\, dt = 1$ *se e só se para algum $x_0 < x_{\sup} \leqslant \infty$,*

$$\overline{F}(x) = c \exp\left(-\int_{x_0}^{x} \frac{g(t)}{a(t)}\, dt\right), \quad x_0 < x < x_{\sup} \qquad (c > 0)$$

onde $\lim\limits_{x \to x_{\sup}} g(x) = 1$ e $a(x)$ é uma função positiva, absolutamente contínua em $]x_0, x_{\sup}[$ com derivada $a'(x)$ verificando $\lim\limits_{x \to x_{\sup}} a'(x) = 0$.

(d) *Para algum $x_0 < x_{\sup} \leqslant \infty$,*

$$\lim_{x \to x_{\sup}} \frac{f(x)}{\left[\overline{F}(x)\right]^2} \int_x^{x_{\sup}} \overline{F}(t)\, dt = 1 \iff t f\big(F^{\leftarrow}(1 - t^{-1})\big) \in \mathcal{R}_0 \iff$$

$$\iff \overline{F}(x) = c \exp\left(-\int_{x_0}^{x} \frac{g(t)}{a(t)}\, dt\right), \quad x_0 < x < x_{\sup} \qquad (c > 0)$$

onde $\lim_{x \to x_{\sup}} g(x) = 1$ e $a(x)$ é uma função positiva, absolutamente contínua em $]x_0, x_{\sup}[$ com derivada $a'(x)$ verificando $\lim\limits_{x \to x_{\sup}} a'(x) = 0$.

Demonstração. Consultar [78], página 64. ∎

Corolário 1.5.4 *Seja F uma função de distribuição absolutamente contínua com extremo direito do suporte $x_{\sup} \leqslant \infty$ e suponhamos que existe algum $x_0 < x_{\sup}$ tal que F possui segunda derivada F'' negativa em $]x_0, x_{\sup}[$.*

(a) *Se*

$$\lim_{x \to x_{\sup}} \frac{\overline{F}(x)\, F''(x)}{\big[F'(x)\big]^2} = -1$$

então $F \in \mathcal{D}(\Lambda)$ *e podemos tomar* $a(x) = \frac{\overline{F}(x)}{F'(x)}$.

(b) *Se* F'' *é não-decrescente,* $F'(x) = -\int_x^{x_{\sup}} F''(t)\, dt$ *e* $F \in \mathcal{D}(\Lambda)$ *então* $\lim\limits_{x \to x_{\sup}} \frac{\overline{F}(x)\, F''(x)}{[F'(x)]^2} = -1$.

(c) $\lim\limits_{x \to x_{\sup}} \frac{\overline{F}(x)\, F''(x)}{[F'(x)]^2} = -1$ *se e só se* F *é uma função de von Mises duas vezes diferenciável.*

Demonstração. Consultar [78], página 66. ∎

O resultado abaixo expõe uma propriedade interessante de algumas distribuições do domínio de atracção maximal da distribuição de Gumbel.

Proposição 1.5.2 *Se* X *é uma v.a. com função de distribuição* $F \in \mathcal{D}(\Lambda)$ *e extremo direito do suporte infinito então* $\overline{F} \in \mathcal{R}_{-\infty}$. *Em particular,* $\mathbb{E}\big[(\mathrm{X}^+)^p\big] < \infty$ *para qualquer* $p > 0$ *onde* $\mathrm{X}^+ = \max(\mathrm{X}, 0)$.

Demonstração. Consultar [26], página 148. ∎

1.6 Análise de regressão

O termo *regressão* foi introduzido, no século XIX, por Francis Galton no seu artigo *Family Likeness in Stature* de 1886. Neste trabalho, Galton estabeleceu que, embora exista uma tendência para que pais altos tenham filhos altos e pais baixos tenham filhos baixos, a média de alturas de filhos nascidos de pais com uma determinada altura (em particular, de pais invulgarmente altos ou baixos) tendem a mover-se ou a "regredir"em direcção à altura média de toda a população. Esta *lei de regressão universal* proposta por Galton, foi confirmada mais tarde por Karl Pearson num artigo de 1903 intitulado *On the Laws of Inheritance* onde foram recolhidos mais de um milhar de registos de alturas de membros de grupos familiares. Mais concretamente, Pearson estabeleceu que a média de alturas de filhos de um grupo de pais altos era inferior à média de alturas dos seus progenitores e que a média de alturas de filhos de um grupo de pais baixos era maior que a média de alturas dos seus progenitores, constatando-se uma igual "regressão"para filhos altos e baixos em direcção à média de todos os indivíduos (nas palavras de Galton esta era a "regressão à mediocridade"). Contudo, a interpretação moderna de regressão é bastante diferente. Em *grosso modo*, a análise de regressão está relacionada com o estudo da dependência de uma variável (variável dependente) a uma ou mais variáveis tendo em vista a estimação e/ou predição da (população) média ou do valor médio da variável dependente à custa dos valores conhecidos ou fixos (em amostras realizadas) das restantes variáveis.

Nas mais diversas ciências, é usual estudar um determinado fenómeno com base em dados (informações resultantes de observações efectuadas ou situações experimentais realizadas), com o objectivo de descrever, explicar ou prever o seu comportamento. Neste processo, é importante saber, ainda que de forma aproximada, o mecanismo subjacente ao fenómeno em estudo com o intuito de estabelecer formulações matemáticas que correspondam fielmente às descrições do mesmo. Estas formulações matemáticas são vulgarmente conhecidas por *modelos*.

Em muitas situações, há apenas interesse em investigar se uma variável quantitativa Y está relacionada com outras variáveis quantitativas $X_1, \ldots, X_p$ e qual a expressão matemática que, porventura, as liga. Assim, de um modo geral, o estudo do comportamento da variável Y como função das outras variáveis $X_1, \ldots, X_p$ pode ser traduzido pelo seguinte modelo,

$$Y = f(X_1, \ldots, X_p; \beta_1, \ldots, \beta_\kappa) \tag{1.6.1}$$

onde $\beta_1, \ldots, \beta_\kappa$ são constantes fixas mas desconhecidas, designadas habitualmente por **parâmetros** do modelo. Na terminologia corrente, a variável Y é conhecida por **variável explicada**, **variável resposta** ou **variável endógena** e as variáveis $X_1, \ldots, X_p$ são chamadas de **variáveis explicativas**, **variáveis controladas**[23] ou **variáveis exógenas**. É especialmente importante o caso em que a relação funcional (1.6.1) assume a forma,

$$Y = \beta_1 X_1 + \ldots + \beta_\kappa X_\kappa \tag{1.6.2}$$

dizendo-se então que estamos perante um **modelo linear** relativamente aos parâmetros.

De seguida, são apresentados alguns exemplos concretos das relações funcionais acima descritas.

Exemplo 1.6.1

O consumo privado, considerado como agregado macroeconómico, constitui uma variável cujo comportamento tem sido amplamente estudado. O modelo mais simples é a **função de consumo keynesiana**, onde o consumo C é explicado a partir do rendimento disponível R, através da relação $C = f(R)$. É habitual propor para f uma função afim $C = \beta_1 + \beta_2 R$, onde β_1 e β_2 são os parâmetros do modelo (β_2 é a propensão marginal para consumir pois mede a variação de C relativa à variação de R satisfazendo $0 < \beta_2 < 1$).

Exemplo 1.6.2

Suponhamos que numa experiência, foram feitas medições da voltagem V relativamente à intensidade de corrente I num determinado circuito eléctrico. Teoricamente, existe uma relação linear entre as variáveis V e I conhecida por **lei de Ohm**: $V = R\,I$ onde R é a *resistência*. Assim, se desejarmos uma estimação de R a partir das medições efectuadas, é razoável considerar a relação $V = \beta I$.

Em muitos casos, pode acontecer que o modelo (1.6.1) não seja linear relativamente aos parâmetros, mas que mediante uma transformação de variável se consiga obter um modelo linear, afirmando-se então que o modelo em causa foi linearizado. Um modelo linear ou linearizável diz-se **intrinsecamente linear** relativamente aos parâmetros.

Exemplo 1.6.3

Considere-se uma unidade produtiva que se dedica ao fabrico de um bem, onde é possível estabelecer (em certas condições) uma relação funcional entre a produção Q do bem e determinada combinação

[23]Note-se que o termo "controladas"usado para designar as variáveis $X_1, \ldots, X_p$ é bastante sugestivo, no sentido em que estas variáveis assumem certos valores fixos no modelo.

de factores produtivos como sucede, por exemplo, com capital K e trabalho L. Esta relação funcional chama-se **função de produção**. Uma especificação muito utilizada é a **função de Cobb-Douglas**,

$$Q = \beta_1 K^{\beta_2} L^{\beta_3} \tag{1.6.3}$$

onde β_1, β_2 e β_3 são constantes positivas (β_2 e β_3 representam as elasticidades da quantidade produzida relativamente ao capital e ao trabalho, respectivamente). A função de Cobb-Douglas é linearizável uma vez que logaritmizando a expressão (1.6.3) obtém-se,

$$\log Q = \log \beta_1 + \beta_2 \log K + \beta_3 \log L.$$

Enfim, outros exemplos de modelos intrinsecamente lineares poderiam ser citados como o **modelo de Michaelis-Menten** que traduz a actividade de algumas enzimas,

$$Y = \frac{\beta_1 X}{\beta_2 + X}.$$

Esta equação relaciona a "velocidade"inicial de uma reacção enzimática com a concentração de substrato X (β_1 indica a velocidade de reacção quando a enzima está completamente saturada com o substrato e a reacção se processa à máxima velocidade e β_2 é a concentração do substrato aquando da metade da velocidade de reacção). Este modelo pode ser transformado num modelo linear expressando o recíproco da velocidade como função da concentração de substrato recíproca,

$$\frac{1}{Y} = \frac{1}{\beta_1} + \frac{\beta_2}{\beta_1}\frac{1}{X}.$$

Naturalmente, pode sempre acontecer que o modelo não seja intrinsecamente linear nos parâmetros. É o que sucede nos próximos exemplos,

Exemplo 1.6.4

A **função de produção CES**,[24]

$$Q = \beta_1 \left[(1-\beta_3)L^{-\beta_4} + \beta_3 K^{-\beta_4}\right]^{-\frac{\beta_2}{\beta_4}}$$

onde $\beta_1 > 0$, $\beta_2 > 0$, $0 < \beta_3 < 1$ e $\beta_4 \in \mathbb{R}$ é não-linear e neste caso, não existe qualquer transformação de Q que permita obter um modelo linear.

Exemplo 1.6.5

No **modelo logístico**,

$$Y = \frac{\beta_1}{1 + \beta_2 e^{-\beta_3 X}}$$

em que $\beta_1, \beta_2, \beta_3 \in \mathbb{R}$ constata-se, sem dificuldade, que estamos perante um modelo intrinsecamente não-linear (para outros exemplos ver [9]).

[24] Função de produção elasticidade de substituição constante.

Por questões de conteúdo em capítulos posteriores, a nossa base de trabalho assentará, salvo indicação em caso contrário, no modelo linear proposto em (1.6.2). Porém, é indispensável avaliar a sua adequação à realidade, nomeadamente para prever a evolução do fenómeno em estudo, sendo pois necessário estimar os parâmetros do modelo. Visto que os resultados que pretendemos desenvolver mais adiante se enquadram exclusivamente num cenário de amostras de grande dimensão, deveremos dispor de:

(a) *Informação* sobre as grandezas (variáveis e parâmetros) do modelo. Uma componente fundamental desta informação consiste nas observações das variáveis, que constituem os dados do modelo;

(b) *Métodos de cálculo* que possibilitem obter estimativas para os parâmetros do modelo, através de um estimador indispensável para o efeito;

(c) Técnicas que permitam analisar a *consistência* do estimador para grandes amostras.

Suponhamos que se têm n observações de cada variável,

$$y_i; x_{i1}, \ldots, x_{i\kappa} \qquad (i = 1, \ldots, n)$$

que constituem os dados do modelo. Convém referir que, os dados se podem classificar em **seccionais** e **temporais**. Os dados são seccionais quando se referem a observações de atributos de determinadas entidades em certo momento ou período de tempo (por exemplo, as quantidades produzidas e as quantidades de factores de produção utilizados referentes às empresas de uma certa indústria num determinado ano, as despesas em bens de consumo e os rendimentos das famílias consideradas numa determinada amostra em determinado ano). Os dados são temporais quando dizem respeito a observações de atributos de uma mesma entidade para vários momentos ou períodos de tempo (por exemplo: as quantidades produzidas e as quantidades de factores de produção utilizados referentes a uma determinada indústria nos últimos 10 anos, o consumo e o rendimento disponível num determinado país nos últimos 20 anos). Para cada observação, o modelo especificado permite escrever a seguinte relação determinística,

$$y_i = \beta_1 x_{i1} + \ldots + \beta_\kappa x_{i\kappa}, \quad i = 1, \ldots, n \tag{1.6.4}$$

i.e. uma função que caracteriza de forma exacta o modo como y_i depende de $x_{i1}, \ldots, x_{i\kappa}$ para cada i. No entanto, facilmente se comprova que, em muitas aplicações (sobretudo em fenómenos sociais), é praticamente impossível dispor de uma relação deste tipo para cada uma das n observações. Por exemplo, admitamos que temos dados relativos ao rendimento X e às despesas de consumo Y de um conjunto de famílias e pretendemos efectuar um orçamento de sobrevivência. Visto que as despesas de consumo familiares dependem, para além do rendimento, de factores tão cruciais como o número de membros e a sua composição, suponhamos ainda que as famílias em estudo possuem o mesmo número de membros e igual composição. Porém, é completamente irrealista esperar que todas as famílias com um dado rendimento X_i mostrem exactamente as mesmas despesas $\mathrm{Y}_i = \beta_1 + \beta_2 \mathrm{X}_i$, pois de entre todas as famílias com o mesmo rendimento existiram variações nas idades, um dos membros pode gostar de carros de luxo, roupas de marca, jogar regularmente em casinos, viajar para o estrangeiro, ... e uma lista sem fim poderia ser aqui anunciada. Por outro lado, alguns destes factores podem nem ser quantificáveis e mesmo que sejam, pode não ser possível obter dados sobre eles. Além disso, muitas variáveis podem sofrer ligeiros efeitos de maneira a que, mesmo com quantidades substanciais de dados, a estimação estatística da sua influência se possa revelar difícil e incerta. Evidentemente, há também que acautelar um certo elemento básico e imprevisível de casualidade na resposta humana.

Esta questão do enquadramento teórico do fenómeno em estudo é ultrapassada recorrendo à "flexibilização"da relação (1.6.4), ou seja, especificando um relacionamento linear expandido que envolva variáveis aleatórias. Deste modo, admite-se que as observações da variável explicada são variáveis aleatórias Y_i e adiciona-se uma variável aleatória e_i à combinação linear das respectivas observações das variáveis explicativas,

$$\mathrm{Y}_i = \beta_1 x_{i1} + \ldots + \beta_\kappa x_{i\kappa} + \mathrm{e}_i, \quad i = 1, \ldots, n. \tag{1.6.5}$$

O termo e_i é designado por **erro aleatório** associado à i-ésima observação e o seu objectivo em (1.6.5) é caracterizar as discrepâncias que emergem entre os valores de Y_i observados e os valores que seriam dados por um relacionamento funcional exacto. Claro está que, o efeito líquido de todas as influências omitidas ou não-mensuráveis do anterior problema do orçamento familiar fica representado por esta variável aleatória. Mais geralmente, na teoria dos erros, os erros aleatórios ou acidentais são devidos a causas diversas e incoerentes, bem como a causas temporais que variam durante observações sucessivas e que escapam a uma análise em função da sua imprevisibilidade, tanto em grandeza como em sentido. Podem ter várias origens como: os instrumentos de medida, pequenas variações das condições ambientais (pressão, temperatura, fontes de ruídos) ou factores relacionados com o próprio observador. Os erros acidentais mudam o resultado em qualquer sentido, aumentando a dispersão de resultados e afectando consequentemente a *precisão* dos mesmos. Obviamente, não podemos predizer qual o valor específico de e_i em cada observação, mas podemos fazer hipóteses sobre as características principais da sua distribuição de probabilidade.

O conjunto das n igualdades (1.6.5) (uma para cada observação) é tradicionalmente conhecido por **modelo de regressão linear**. Utilizando a notação matricial, podemos escrever o modelo de regressão linear sinteticamente,

$$\mathbf{Y} = \boldsymbol{X}\boldsymbol{\beta} + \mathbf{e} \tag{1.6.6}$$

onde,

$$\mathbf{Y} = \begin{bmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ \vdots \\ \mathrm{Y}_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1\kappa} \\ x_{21} & x_{22} & \ldots & x_{2\kappa} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{n\kappa} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_\kappa \end{bmatrix} \quad \text{e} \quad \mathbf{e} = \begin{bmatrix} \mathrm{e}_1 \\ \mathrm{e}_2 \\ \vdots \\ \mathrm{e}_n \end{bmatrix}$$

são respectivamente, o **vector das observações** (aleatórias) da variável explicada, a **matriz de modelo**, o **vector parâmetro** e o **vector dos erros** (aleatórios). Num quadro mais generalista, que possibilite a inclusão de funções não-lineares nos parâmetros, podemos idealizar o **modelo de regressão não-linear**,

$$\begin{cases} \mathrm{Y}_1 = f(\boldsymbol{x}_{1\bullet}, \boldsymbol{\beta}) + \mathrm{e}_1 \\ \mathrm{Y}_2 = f(\boldsymbol{x}_{2\bullet}, \boldsymbol{\beta}) + \mathrm{e}_2 \\ \qquad\vdots \\ \mathrm{Y}_n = f(\boldsymbol{x}_{n\bullet}, \boldsymbol{\beta}) + \mathrm{e}_n \end{cases}$$

em que $f\colon \mathbb{R}^\kappa \times \Theta \to \mathbb{R}$ é uma função conhecida à partida e

$$\begin{aligned} \boldsymbol{x}_{1\bullet} &= (x_{11}, x_{12}, \ldots, x_{1\kappa}) \\ \boldsymbol{x}_{2\bullet} &= (x_{21}, x_{22}, \ldots, x_{2\kappa}) \\ &\vdots \\ \boldsymbol{x}_{n\bullet} &= (x_{n1}, x_{n2}, \ldots, x_{n\kappa}) \end{aligned}$$

são os vectores das variáveis explicativas ou exógenas (ver [9] ou [33]).

Apresentado o modelo de regressão linear (1.6.6) que analisaremos detalhadamente nos próximos capítulos com hipóteses adicionais sobre os coeficientes da matriz de modelo e as componentes do vector dos erros, avançamos para o método de cálculo que usaremos para estimar os respectivos parâmetros do modelo. Por questões técnicas, assumiremos que $n \geqslant \kappa$ e também que os vectores das n observações de cada uma das κ variáveis explicadas (endógenas),

$$\begin{aligned} \boldsymbol{x}_{\bullet 1} &= (x_{11}, x_{21}, \ldots, x_{n1}) \\ \boldsymbol{x}_{\bullet 2} &= (x_{12}, x_{22}, \ldots, x_{n2}) \\ &\vdots \\ \boldsymbol{x}_{\bullet \kappa} &= (x_{1\kappa}, x_{2\kappa}, \ldots, x_{n\kappa}) \end{aligned}$$

são linearmente independentes, i.e. admitiremos que a característica da matriz de modelo $\boldsymbol{X}$ é κ. Por outras palavras, recordando que existe **multicolinearidade exacta** da matriz de modelo $\boldsymbol{X}$ quando a sua característica é inferior a κ, trabalharemos sempre sob a ausência de multicolinearidade exacta da matriz de modelo $\boldsymbol{X}$. De referir ainda que, excluiremos também a hipótese de **multicolinearidade quasi-exacta** da matriz de modelo, ou seja, evitaremos situações em que a característica de $\boldsymbol{X}$ é κ mas os vectores $\boldsymbol{x}_{\bullet 1}, \boldsymbol{x}_{\bullet 2}, \ldots, \boldsymbol{x}_{\bullet \kappa}$ são "quase"linearmente dependentes, o que acontece se $\boldsymbol{x}_{\bullet 1}$ estiver "perto"de ser linearmente gerado por $\boldsymbol{x}_{\bullet 2}, \ldots, \boldsymbol{x}_{\bullet \kappa}$. Por exemplo, considerando

$$\boldsymbol{X}\boldsymbol{c} = c_1\boldsymbol{x}_{\bullet 1} + c_2\boldsymbol{x}_{\bullet 2} + \ldots + c_\kappa\boldsymbol{x}_{\bullet \kappa} \doteq \boldsymbol{0}$$

onde

$$\boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_\kappa \end{bmatrix}$$

com alguma componente não-nula e "$\doteq$"significa "quase igual a". Se $c_1 \neq 0$ (se $c_1 = 0$ rearrange-se as colunas da matriz de modelo $\boldsymbol{X}$ de forma a não ser zero) então,

$$\boldsymbol{x}_{\bullet 1} \doteq \left(-\frac{c_2}{c_1}\right)\boldsymbol{x}_{\bullet 2} + \ldots + \left(-\frac{c_\kappa}{c_1}\right)\boldsymbol{x}_{\bullet \kappa} \tag{1.6.7}$$

ou

$$\boldsymbol{x}_{\bullet 1} = \left(-\frac{c_2}{c_1}\right)\boldsymbol{x}_{\bullet 2} + \ldots + \left(-\frac{c_\kappa}{c_1}\right)\boldsymbol{x}_{\bullet \kappa} + \boldsymbol{v}_1$$

em que $\boldsymbol{v}_1$ é precisamente a diferença entre os membros esquerdo e direito de (1.6.7). Nos fenómenos económicos, é usual encontrar o problema da multicolinearidade uma vez que as observações de variáveis económicas diferentes (particularmente as referentes a dados agregados) estão fortemente correlacionadas, o que dá lugar a "quase"dependências lineares.

Nestas condições, podemos garantir a existência da matriz $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$ facto que, como se vai ver de seguida, terá uma importância decisiva para a estimação dos parâmetros do modelo de regressão linear. De notar ainda que a diferença $n - \kappa$ constitui o número de **graus de liberdade** do modelo.

O método que iremos utilizar para estimar o vector parâmetro $\boldsymbol{\beta}$ (e portanto os parâmetros) do modelo de regressão linear (1.6.6) designa-se por **método dos mínimos quadrados**. Convém sublinhar que o vector parâmetro $\boldsymbol{\beta}$ assume valores num dado subconjunto Θ de $\mathbb{R}^\kappa$ chamado **espaço parâmetro**. A título ilustrativo,

$$\Theta = \left\{(\beta_1, \beta_2) \in \mathbb{R}^2 \colon\ 0 < \beta_2 < 1\right\}$$

é o espaço parâmetro associado ao Exemplo 1.6.1.

Definição 1.6.1 O **estimador dos mínimos quadrados (EMQ)** do vector parâmetro $\boldsymbol{\beta}$ é o vector $\widetilde{\boldsymbol{\beta}}$ do espaço parâmetro $\Theta \subset \mathbb{R}^\kappa$ que minimiza a função $u \colon \Theta \to \mathbb{R}$ definida por,

$$u(\boldsymbol{\beta}) = \left(\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta}\right)^T\left(\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta}\right).$$

Observação 1.6.1 Observe-se que o método dos mínimos quadrados propõe o seguinte critério: escolher $\boldsymbol{\beta}$ no espaço parâmetro de forma a minimizar a soma do quadrado dos erros,

$$e_1^2 + e_2^2 + \ldots + e_n^2 \quad \text{ou} \quad \mathbf{e}^T\mathbf{e}.$$

Retomando a expressão de u da Definição 1.6.1 e utilizando propriedades bem conhecidas do cálculo matricial obtém-se sucessivamente,

$$\begin{aligned}
u(\boldsymbol{\beta}) &= \mathbf{Y}^T\mathbf{Y} - \boldsymbol{\beta}^T\boldsymbol{X}^T\mathbf{Y} - \mathbf{Y}^T\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} \\
&= \mathbf{Y}^T\mathbf{Y} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\mathbf{Y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} \\
&= \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{Y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} + \mathbf{Y}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{Y} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\mathbf{Y} \\
&= \mathbf{Y}^T\left[\boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]\mathbf{Y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} + \\
&\qquad + \mathbf{Y}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{Y} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{Y} \\
&= \mathbf{Y}^T\left[\boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\right]\mathbf{Y} + \left[\boldsymbol{\beta} - \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{Y}\right]^T\boldsymbol{X}^T\boldsymbol{X}\left[\boldsymbol{\beta} - \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{Y}\right]
\end{aligned}$$

donde para encontrarmos o minimizante $\widetilde{\boldsymbol{\beta}}$ basta ter em conta que,

$$\left[\boldsymbol{\beta} - \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{Y}\right]^T\boldsymbol{X}^T\boldsymbol{X}\left[\boldsymbol{\beta} - \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{Y}\right] \tag{1.6.8}$$

é uma forma quadrática definida positiva de matriz $\boldsymbol{X}^T\boldsymbol{X}$. Consequentemente, o valor que minimiza a forma quadrática (1.6.8) será aquele que a anula o que acontece se,

$$\widetilde{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{Y}. \tag{1.6.9}$$

A solução (1.6.9) obtida designa-se habitualmente por **solução dos mínimos quadrados** e é a solução exacta e única do sistema de Cramer $\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^T\mathbf{Y}$ conhecido por **sistema de equações normais** dos mínimos quadrados.

Observação 1.6.2 Convém destacar que existem outros métodos de cálculo que possibilitam a estimação dos parâmetros em cenários mais gerais, nomeadamente, quando o problema da multicolinearidade exacta se coloca e não existe a inversa da matriz $\boldsymbol{X}^T\boldsymbol{X}$ (pois a matriz de modelo tem característica inferior a κ), não podendo a solução dos mínimos quadrados ser alcançada. Uma dessas técnicas de cálculo chama-se **método das componentes principais** que, no entanto, não será abordada no presente texto.

Encontrado que está o EMQ, resta-nos agora definir formalmente o conceito de consistência. Antes porém, estabeleçamos algumas noções elementares de teoria da amostragem.

Em Estatística, o termo **população** designa um conjunto de elementos cujos atributos são objecto de um determinado estudo. Para conhecer de forma completa uma população tem de analisar-se todos os seus elementos, isto é, realizar um censo ou indagação completa. Exceptuando os casos em que a população tem dimensão modesta e é acessível, raramente é possível analisar todos os elementos de uma população finita e, evidentemente, é sempre impossível observar todos os elementos de uma população infinita. Portanto, o estudo dos atributos de uma população tem de ser feito sobre um seu subconjunto finito, que se designa por **amostra**. Os termos amostra e população referem-se a conjuntos de unidades estatísticas contudo, por abuso de linguagem, estes termos também são utilizados para identificar os conjuntos de valores assumidos pelo atributo em estudo.

Quando um observador, numa perspectiva inferencial, dispõe de uma amostra de n observações $(x_1, \ldots, x_n)$, supõe que tal amostra é uma realização do vector aleatório

$$(\mathrm{X}_1, \ldots, \mathrm{X}_n).$$

Este vector designa-se habitualmente por **amostra aleatória** de dimensão n. Um processo de amostragem, ou seja, o processo seguido para escolher os elementos da população a incluir na amostra, dá lugar a $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ e conduz virtualmente a muitas amostras diferentes. O subconjunto de $\mathbb{R}^n$ de todas as amostras passíveis de serem seleccionadas chama-se **espaço amostra**. Existem vários métodos de selecção de amostras aleatórias que condicionam a forma de fazer inferências a partir das mesmas. Um processo particular de amostragem, a **amostragem casual**, acontece quando as n variáveis aleatórias observadas, componentes do vector $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ são independentes e identicamente distribuídas.

Definição 1.6.2 Uma **estatística** é um vector aleatório $\mathbf{T}(\mathrm{X}_1, \ldots, \mathrm{X}_n)$, função da amostra aleatória $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$, que não envolve qualquer parâmetro desconhecido.

Dada uma amostra aleatória $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ de uma população com função de distribuição conjunta dependendo, entre possivelmente outros, de um vector parâmetro $\boldsymbol{\theta}$ (desconhecido) pertencente ao espaço parâmetro $\Theta \subset \mathbb{R}^\kappa$ então as estatísticas que são utilizadas com o propósito de estimar $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_\kappa)$ denotam-se por $\mathbf{T}(\mathrm{X}_1, \ldots, \mathrm{X}_n)$, $\mathbf{T}'(\mathrm{X}_1, \ldots, \mathrm{X}_n)$, $\mathbf{T}''(\mathrm{X}_1, \ldots, \mathrm{X}_n)$,... . Assim, a qualquer variável aleatória função da amostra aleatória $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ chamamos **estimador** e representamos por $\mathbf{T}_n = \mathbf{T}(\mathrm{X}_1, \ldots, \mathrm{X}_n)$.

Definição 1.6.3 Um estimador $\mathbf{T}_n = \mathbf{T}(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ diz-se:

(F1) **fortemente consistente** de $\boldsymbol{\theta}$ se $\mathbf{T}_n \xrightarrow{\text{q.c.}} \boldsymbol{\theta}$ para cada $\boldsymbol{\theta} \in \Theta$ fixo.

(f2) **fracamente consistente** de $\boldsymbol{\theta}$ se $\mathbf{T}_n \xrightarrow{\mathbb{P}} \boldsymbol{\theta}$ para cada $\boldsymbol{\theta} \in \Theta$ fixo.

Definição 1.6.4 Dado $p > 0$ diz-se que o estimador $\mathbf{T}_n = \mathbf{T}_n(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ é **consistente em momento de ordem $\boldsymbol{p}$** de $\boldsymbol{\theta}$ se $\mathbf{T}_n \xrightarrow{\mathscr{L}_p} \boldsymbol{\theta}$ para cada $\boldsymbol{\theta} \in \Theta$ fixo. A consistência em momento de ordem 2 é chamada **consistência em média quadrática**.

A análise da consistência forte do EMQ nos mais diversos cenários será, agora, o tema dos próximos capítulos...

Capítulo 2

A consistência forte

2.1 Condições gerais de suficiência

Consideremos o modelo de regressão linear (1.6.6) reescrito do seguinte modo,

$$\mathrm{Y}_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \mathrm{e}_i, \quad 1 \leqslant i \leqslant n, \quad n \geqslant 1 \tag{2.1.1}$$

onde

$$\boldsymbol{x}_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{i\kappa} \end{bmatrix}, \qquad i = 1, \ldots, n$$

é um vector coluna não-aleatório com κ componentes, e_i é um erro aleatório relativo à i-ésima observação ($i = 1, \ldots, n$) e o $\boldsymbol{\beta}$ o vector parâmetro.

Desde a década de sessenta, a consistência do EMQ em modelos lineares tem atraído a atenção de muitos estatísticos em todo o mundo. Sensivelmente, na segunda metade da década de setenta este problema foi satisfatoriamente resolvido para o caso em que os erros aleatórios possuiam variância finita. Mais concretamente, nos anos de 1978 (ver [53]) e 1979 (ver [54]) Lai, Robbins e Wei mostraram que se a sucessão $\{\mathrm{e}_i,\, i \geqslant 1\}$ fosse i.i.d. com $\mathbb{E}(\mathrm{e}_1) = 0$ e $0 < \mathbb{E}(\mathrm{e}_1^2) < \infty$ então uma condição suficiente para a consistência forte de $\widetilde{\boldsymbol{\beta}}$ seria,

$$\boldsymbol{S}_n^{-1} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \longrightarrow \mathbf{O} \tag{2.1.2}$$

quando $n \to \infty$. Contudo, em 1976, Drygas já tinha demonstrado, com as mesmas hipóteses para sucessão de erros, que a condição (2.1.2) era necessária para se obter a consistência forte do EMQ. Portanto, ficou estabelecido que se a sucessão $\{\mathrm{e}_i,\, i \geqslant 1\}$ for i.i.d. com $\mathbb{E}(\mathrm{e}_1) = 0$ e $0 < \mathbb{E}(\mathrm{e}_1^2) < \infty$ então a condição (2.1.2) é necessária e suficiente para que $\widetilde{\boldsymbol{\beta}}$ seja fortemente consistente.

Posteriormente, durante a década de oitenta, alguns autores asiáticos estudaram este problema a fundo sob hipóteses mais fracas para a sucessão de erros aleatórios, nomeadamente, quando os erros apenas possuem momentos absolutos finitos de ordem inferior a 2. Inicialmente, as investigações foram influenciadas pela crença da condição $\boldsymbol{S}_n^{-1} \longrightarrow \mathbf{O}$ $(n \to \infty)$ ter um papel crucial na consistência de $\widetilde{\boldsymbol{\beta}}$. Mas mais tarde constatou-se que esta condição, apesar de ser essencial para o caso em que a variância dos erros é finita, é insuficiente quando os erros possuem momentos absolutos finitos de ordem inferior a 2. Aliás, neste cenário, torna-se mesmo impossível obter simultaneamente condições necessárias e suficientes para a consistência forte conseguindo-se apenas condições de suficiência. Já na década de noventa, com o objectivo de ultrapassar este irremediável obstáculo, os autores Xiru Chen e Mingzhong Jin estabeleceram um conceito mais geral de "condição necessária e suficiente"para a consistência forte de $\widetilde{\boldsymbol{\beta}}$ chamada *condição necessária e suficiente de primeira espécie* em oposição à tradicional noção de "condição necessária e suficiente"(designada então de *condição necessária e suficiente de segunda espécie*). Evidentemente, este novo conceito surgiu com o propósito de dar sentido à não existência de condições necessárias no cenário desejado. Na realidade, o conceito usual de "condição necessária e suficiente", apesar de ser mais forte do que o novo conceito estabelecido, goza da inconveniência de não existir nalgumas situações; por outro lado, o conceito de *condição necessária e suficiente de primeira espécie* tem a vantagem de existir sempre ficando, sob um certo ponto de vista, garantida a existência de condições necessárias e suficientes (ver [15] e [16] para mais detalhes).

Este estudo acabou por ter um enorme significado pois, devido à sua natureza probabilística associada às somas de v.a.'s independentes, o problema da consistência forte pode ser encarado como uma extensão natural das leis de grandes números. Adicionalmente, através da análise do comportamento do EMQ sob hipóteses mais fracas para os momentos absolutos, são reveladas novas propriedades do EMQ que, curiosamente, contrastam com a situação em que a variância dos erros é finita tais como:

1. O facto da equivalência da consistência fraca e forte do EMQ ser verdadeira quando a sucessão $\{\mathrm{e}_i,\, i \geqslant 1\}$ é i.i.d. com $\mathbb{E}\left(\mathrm{e}_1^2\right) < \infty$ e falsa no caso dos e_i's possuírem apenas momentos absolutos finitos de ordem $r < 2$.

2. É do conhecimento geral que, quando uma sucessão de erros $\{\mathrm{e}_i,\, i \geqslant 1\}$ é i.i.d. com $\mathbb{E}\left(\mathrm{e}_1^2\right) < \infty$ então a introdução de algum parâmetro linear (incómodo) no modelo (2.1.1) não altera o EMQ de não-consistente para consistente. Sob condições mais fracas nos momentos absolutos esta propriedade já não é válida.

O objectivo principal desta secção será estender as condições de suficiência desenvolvidas por Xiru Chen e Mingzhong Jin à situação em que a sucessão $\{\mathrm{e}_i,\, i \geqslant 1\}$ é i.i.d. com $\mathbb{E}\,|\mathrm{e}_1|^r < \infty$ para algum $0 < r \leqslant 1$.

Suponhamos que a matriz $\boldsymbol{S}_i$ é invertível para valores grandes de i e escreva-se,

$$\boldsymbol{d}_i = \boldsymbol{S}_i^{-1}\boldsymbol{x}_i, \quad i \geqslant 1$$

com $\boldsymbol{d}_i$ definido de forma arbitrária para valores pequenos de i (quando a matriz $\boldsymbol{S}_i^{-1}$ não existe). Defina-se ainda $N(K) = \#\left\{i \geqslant 1\colon\ \|\boldsymbol{d}_i\| \geqslant K^{-1}\right\}$.

Recuperando a matriz $\boldsymbol{X}_n = \big[x_{ij}\big]_{\substack{i=1,\dots,n \\ j=1,\dots,\kappa}}$ do modelo de regressão linear (1.6.6) temos,[1]

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \big(\boldsymbol{X}_n^T \boldsymbol{X}_n\big)^{-1} \boldsymbol{X}_n^T \mathbf{e} \tag{2.1.3}$$

por substituição do vector das observações $\mathbf{Y} = \boldsymbol{X}_n \boldsymbol{\beta} + \mathbf{e}$ em (1.6.9). Observando que $\big(\boldsymbol{X}_n^T \boldsymbol{X}_n\big)^{-1} = \boldsymbol{S}_n^{-1}$ (para valores grandes de n) somos conduzidos à seguinte expressão do EMQ,

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \mathrm{e}_i \tag{2.1.4}$$

pelo que $\widetilde{\boldsymbol{\beta}}$ é fortemente consistente se e só se $\boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \mathrm{e}_i \xrightarrow{\text{q.c.}} \mathbf{0}$.

Abrimos os resultados sobre a consistência forte do EMQ com um teorema que foi anunciado em [63] e que constitui uma extensão da lei forte de Kolmogorov para a regressão múltipla.

Teorema 2.1.1 *Seja* $\{\mathrm{e}_i,\ i \geqslant 1\}$ *uma sucessão de erros i.i.d. do modelo de regressão linear* (2.1.1) *com* $\mathbb{E}(\mathrm{e}_1) = 0$ *e suponhamos que,*

(a) $\displaystyle\lim_{n\to\infty} \frac{1}{n} \boldsymbol{S}_n = \boldsymbol{H}$ *com* $\boldsymbol{H}$ *definida positiva.*

(b) *As linhas da matriz de modelo* $\boldsymbol{X}_n$ *pertencem a um compacto de* $\mathbb{R}^\kappa$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Consultar [63], página 462. ∎

Teorema 2.1.2 *Seja* $\{\mathrm{e}_i,\ i \geqslant 1\}$ *uma sucessão de erros i.i.d. do modelo de regressão linear* (2.1.1) *com* $\mathbb{E}(\mathrm{e}_1) = 0$ *e* $\mathbb{E}\,|\mathrm{e}_1|^r < \infty$ *para algum* $1 < r < 2$. *Se,*

(a) $\displaystyle\lim_{n\to\infty} \boldsymbol{S}_n^{-1} = \mathbf{O}$ *e* $\displaystyle\limsup_{n\to\infty} \sum_{i=1}^{n} \big\|\boldsymbol{S}_n^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^T\big\| < \infty$.

(b) $N(K) = O\big(K^r\big)$ *quando* $K \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. A demonstração da convergência quase certa da série $\sum_{i=1}^{\infty} d_i \mathrm{e}_i$, onde os d_i's representam uma das j-ésimas componentes do vector $\boldsymbol{d}_i = \boldsymbol{S}_i^{-1} \boldsymbol{x}_i$, pode ser vista em [17] ou [43]. A conclusão segue da aplicação do lema de Kronecker matricial com $\boldsymbol{A}_i = \boldsymbol{S}_i$ e $\boldsymbol{v}_i = \boldsymbol{d}_i \mathrm{e}_i$. ∎

Teorema 2.1.3 *Seja* $\{\mathrm{e}_i,\ i \geqslant 1\}$ *uma sucessão de erros i.i.d. do modelo de regressão linear* (2.1.1) *com* $\mathbb{E}\,|\mathrm{e}_1|^r < \infty$ *para algum* $0 < r \leqslant 1$. *Se,*

(a) $\displaystyle\lim_{n\to\infty} \boldsymbol{S}_n^{-1} = \mathbf{O}$ *e* $\displaystyle\limsup_{n\to\infty} \sum_{i=1}^{n} \big\|\boldsymbol{S}_n^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^T\big\| < \infty$.

[1] O índice n em $\boldsymbol{X}_n$ relativo ao número de linhas da matriz de modelo $\big[x_{ij}\big]_{\substack{i=1,\dots,n \\ j=1,\dots,\kappa}}$ é destacado porque vão acontecer passagens ao limite em n.

(b) $N(K) = O\left(\dfrac{K^r}{\log^s K}\right)$ *quando* $K \to \infty$ *para qualquer* $s > 1$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Seja d_i uma das j-ésimas componentes de $\boldsymbol{d}_i = \boldsymbol{S}_i^{-1}\boldsymbol{x}_i$ e $\mathrm{e}_i' = \mathrm{e}_i I_{\{|\mathrm{e}_i d_i| \leqslant 1\}}$. Demonstremos que,

$$\sum_{i=1}^{\infty} d_i \mathrm{e}_i \quad \text{converge q.c.}$$

utilizando o teorema das três séries de Kolmogorov:

i. $\displaystyle\sum_{i=1}^{\infty} \mathbb{P}\big(|d_i \mathrm{e}_i| > 1\big) < \infty.$

Uma vez que,

$$\mathbb{P}\big(|d_i\mathrm{e}_i| > 1\big) \leqslant \mathbb{P}\left(|d_i\mathrm{e}_i|^r I_{\{|\mathrm{e}_i d_i|>1\}} > 1\right) \leqslant |d_i|^r\, \mathbb{E}\left(|\mathrm{e}_i|^r I_{\{|\mathrm{e}_i d_i|>1\}}\right)$$

vamos analisar a série,

$$\sum_{i=1}^{\infty} |d_i|^r\, \mathbb{E}\big(|\mathrm{e}_i|^r I_{\{|\mathrm{e}_i d_i|>1\}}\big).$$

Considerando,

$$p_i = \mathbb{P}\big(i-1 \leqslant |\mathrm{e}_1| < i\big), \quad i \geqslant 1 \tag{2.1.5}$$

$$K-1 \leqslant |d_i|^{-1} < K, \quad K \geqslant 1 \tag{2.1.6}$$

vem,

$$|d_i|^r\, \mathbb{E}\left(|\mathrm{e}_i|^r I_{\{|\mathrm{e}_i d_i|>1\}}\right) \leqslant (K-1)^{-r} \sum_{j=K}^{\infty} j^r p_j, \quad K \geqslant 2$$

já que,

$$|\mathrm{e}_i|^r I_{\{|\mathrm{e}_i d_i|>1\}} \leqslant |\mathrm{e}_i|^r I_{\{|\mathrm{e}_i|>K-1\}} \leqslant \sum_{j=K}^{\infty} |\mathrm{e}_i|^r\, I_{\{j-1 \leqslant |\mathrm{e}_i| < j\}}$$

implica,

$$\mathbb{E}\left(|\mathrm{e}_i|^r I_{\{|\mathrm{e}_i d_i|>1\}}\right) \leqslant \sum_{j=K}^{\infty} j^r p_j.$$

Escrevendo $\widetilde{N}(K) = \#\left\{i \geqslant 1 :\ |d_i|^{-1} \leqslant K\right\}$ obtém-se,[2]

$$\#\left\{i \geqslant 1 :\ K-1 < |d_i|^{-1} \leqslant K\right\} = \widetilde{N}(K) - \widetilde{N}(K-1)$$

[2]Observe-se que $\widetilde{N}(K) \leqslant N(K)$ pois $\{i \geqslant 1 :\ |d_i|^{-1} \leqslant K\} \subset \{i \geqslant 1 :\ \|\boldsymbol{d}_i\| \geqslant K^{-1}\}$.

donde,

$$\sum_{i=1}^{\infty} |d_i|^r \,\mathbb{E}\big(\, |\mathrm{e}_i|^r \, I_{\{|\mathrm{e}_i d_i|>1\}}\big) \leqslant \widetilde{N}(1) \sup_{i\geqslant 1} |d_i|^r \,\mathbb{E}\, |\mathrm{e}_i|^r + \\ + \sum_{K=2}^{\infty} \sum_{\{i:\ K-1<|d_i|^{-1}\leqslant K\}} (K-1)^{-r} \sum_{j=K}^{\infty} j^r \, p_j$$

i.e.

$$\sum_{i=1}^{\infty} |d_i|^r \,\mathbb{E}\big(\, |\mathrm{e}_i|^r \, I_{\{|\mathrm{e}_i d_i|>1\}}\big) \leqslant \widetilde{N}(1) \sup_{i\geqslant 1} |d_i|^r \,\mathbb{E}\, |\mathrm{e}_i|^r + \\ + \sum_{K=2}^{\infty} \Big(\widetilde{N}(K) - \widetilde{N}(K-1)\Big) (K-1)^{-r} \sum_{j=K}^{\infty} j^r \, p_j.$$

Mas,

$$\sum_{K=2}^{\infty} \Big(\widetilde{N}(K) - \widetilde{N}(K-1)\Big)(K-1)^{-r} \sum_{j=K}^{\infty} j^r \, p_j = \sum_{j=1}^{\infty} \widetilde{N}(j+1) p_j + \\ + \sum_{j=2}^{\infty} \left(\sum_{K=2}^{j} \widetilde{N}(K)\Big((K-1)^{-r} - K^{-r}\Big) \right) j^r \, p_j - \sum_{j=1}^{\infty} \widetilde{N}(1) j^r \, p_j$$

e visto que $\widetilde{N}(K) = O\left(\frac{K^r}{\log^s K}\right)$ com $s > 1$ tem-se,

$$\widetilde{N}(K)\Big((K-1)^{-r} - K^{-r}\Big) = O\left(\frac{1}{K \log^s K}\right)$$

atendendo ao facto de,

$$0 \leqslant (K-1)^{-r} - K^{-r} = r \int_{K-1}^{K} t^{-r-1} \, dt \leqslant r(K-1)^{-r-1} \leqslant \\ \leqslant 2^{r+1} r \cdot K^{-r-1}, \quad \forall K \geqslant 2.$$

Então para alguma constante $c_1(r,s) > 0$,

$$\sum_{j=2}^{\infty} \left(\sum_{K=2}^{j} \widetilde{N}(K)\left((K-1)^{-r} - K^{-r}\right) \right) j^r \, p_j \leqslant \sum_{j=2}^{\infty} c_1(r,s) j^r \, p_j$$

pois,

$$0 \leqslant \sum_{K=2}^{j} \frac{1}{K \log^s K} \leqslant \frac{1}{2\log^s 2} + \int_2^j \frac{1}{K \log^s K} \, dt \leqslant \frac{1}{2 \log^s 2} + \frac{\log^{1-s} 2}{s-1}.$$

As séries $\sum_{j=2}^{\infty} c_1(r,s) j^r p_j$, $\sum_{j=1}^{\infty} \widetilde{N}(j+1) p_j$ e $\sum_{j=1}^{\infty} \widetilde{N}(1) j^r p_j$ são convergentes por comparação com $\sum_{j=1}^{\infty} (j-1)^r p_j \leqslant \mathbb{E}\, |\mathrm{e}_1|^r < \infty$ e o termo

$$\widetilde{N}(1) \sup_{i\geqslant 1} |d_i|^r \,\mathbb{E}\, |\mathrm{e}_1|^r$$

é limitado porque $d_i \longrightarrow 0$ quando $i \to \infty$ já que, por hipótese, $\limsup_{i\to\infty} \left\| \boldsymbol{S}_i^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T \right\| < \infty$. Consequentemente,

$$\sum_{i=1}^{\infty} \mathbb{P}\big(\,|d_i \mathrm{e}_i| \geqslant 1\big) < \infty.$$

ii. $\displaystyle\sum_{i=1}^{\infty} \left|\mathbb{E}\big(d_i \mathrm{e}_i'\big)\right| < \infty.$

De (2.1.5) e (2.1.6) obtém-se,

$$\left|\mathbb{E}\big(d_i \mathrm{e}_i'\big)\right| \leqslant |d_i|\, \mathbb{E}\big(\,|\mathrm{e}_i|\, I_{\{|\mathrm{e}_i d_i| \leqslant 1\}}\big) \leqslant (K-1)^{-1} \sum_{j=1}^{K} j\, p_j$$

porque,

$$|\mathrm{e}_i|\, I_{\{|\mathrm{e}_i d_i|\leqslant 1\}} \leqslant |\mathrm{e}_i|\, I_{\{|\mathrm{e}_i|<K\}} \leqslant \sum_{j=1}^{K} |\mathrm{e}_i|\, I_{\{j-1\leqslant|\mathrm{e}_i|<j\}} \leqslant \sum_{j=1}^{K} j\, I_{\{j-1\leqslant|\mathrm{e}_i|<j\}}.$$

Então,

$$\sum_{i=1}^{\infty} \left|\mathbb{E}\big(d_i \mathrm{e}_i'\big)\right| \leqslant p_1 \widetilde{N}(1) \sup_{i\geqslant 1} |d_i| + \sum_{K=2}^{\infty} \sum_{\left\{i:\ K-1<|d_i|^{-1}\leqslant K\right\}} (K-1)^{-1} \sum_{j=1}^{K} j\, p_j$$

ou seja,

$$\sum_{i=1}^{\infty} \left|\mathbb{E}\big(d_i \mathrm{e}_i'\big)\right| \leqslant \widetilde{N}(1) \sup_{i\geqslant 1} |d_i| + \sum_{K=2}^{\infty} \Big(\widetilde{N}(K) - \widetilde{N}(K-1)\Big)\, (K-1)^{-1} \sum_{j=1}^{K} j\, p_j$$

onde,

$$\begin{aligned}\sum_{K=2}^{\infty} \Big(\widetilde{N}(K) - \widetilde{N}(K-1)\Big)(K-1)^{-1} \sum_{j=1}^{K} j\, p_j &= -p_1\widetilde{N}(1) - \sum_{j=2}^{\infty} \widetilde{N}(j-1) \left(\frac{j}{j-1}\right) p_j + \\ &+ \sum_{j=2}^{\infty} p_1 \widetilde{N}(K)\Big((K-1)^{-1} - K^{-1}\Big) + \sum_{j=2}^{\infty} \left(\sum_{K=j}^{\infty} \widetilde{N}(K)\Big((K-1)^{-1} - K^{-1}\Big)\right) j\, p_j.\end{aligned}$$

Visto que $\widetilde{N}(K) = O\left(\frac{K^r}{\log^s K}\right)$ com $s > 1$ tem-se,

$$\widetilde{N}(K)\Big((K-1)^{-1} - K^{-1}\Big) = O\left(\frac{1}{K^{2-r}\log^s K}\right).$$

Quando $0 < r < 1$ vem,

$$\begin{aligned}\sum_{K=j}^{\infty} \frac{1}{K^{2-r}\log^s K} \leqslant \frac{1}{\log^s 2} \sum_{K=j}^{\infty} \frac{1}{K^{2-r}} &\leqslant \frac{1}{\log^s 2} \left(j^{r-2} + \int_j^{\infty} \frac{1}{t^{2-r}}\, dt \right) \leqslant \\ &\leqslant \frac{1}{\log^s 2} \left(j^{r-1} + \frac{j^{r-1}}{1-r} \right) \leqslant \frac{j^{r-1}}{\log^s 2} \left(\frac{2-r}{1-r} \right).\end{aligned} \tag{2.1.7}$$

para $j \geqslant 2$ o que implica,

$$\sum_{j=2}^{\infty}\left(\sum_{K=j}^{\infty}\widetilde{N}(K)\Big((K-1)^{-1}-K^{-1}\Big)\right)j\,p_j \leqslant \sum_{j=2}^{\infty} c_2(r,s)j^r p_j$$

para alguma constante $c_2(r,s)>0$. Se $r=1$ então,

$$\sum_{K=j}^{\infty}\frac{1}{K\log^s K} \leqslant \frac{1}{j\log^s j}+\int_j^{\infty}\frac{1}{t\log^s t}\,dt = \frac{1}{j\log^s j}+\frac{\log^{1-s} j}{s-1} \tag{2.1.8}$$

e

$$\sum_{j=2}^{\infty}\left(\sum_{K=j}^{\infty}\widetilde{N}(K)\Big((K-1)^{-1}-K^{-1}\Big)\right)j\,p_j \leqslant \sum_{j=2}^{\infty} c_3\left(\frac{1}{j\log^s j}+\frac{\log^{1-s} j}{(s-1)}\right)jp_j$$

para alguma constante $c_3>0$. De (2.1.7) e (2.1.8) a convergência da série

$$\sum_{j=2}^{\infty}\widetilde{N}(K)\Big((K-1)^{-1}-K^{-1}\Big).$$

fica assegurada. As séries $\sum_{j=2}^{\infty} c_2(r,s)j^r p_j$, $\sum_{j=1}^{\infty}\widetilde{N}(j-1)\left(\frac{j}{j-1}\right)p_j$ e

$$\sum_{j=2}^{\infty} c_3\left(\frac{1}{j\log^s j}+\frac{\log^{1-s} j}{(s-1)}\right)jp_j$$

também são convergentes por comparação com a série $\sum_{j=1}^{\infty}(j-1)^r p_j<\infty$. Portanto,

$$\sum_{i=1}^{\infty}\left|\mathbb{E}\big(d_i \mathrm{e}'_i\big)\right| < \infty.$$

iii. $\displaystyle\sum_{i=1}^{\infty}\mathbb{E}\left(d_i^2 {\mathrm{e}'_i}^2\right)<\infty.$

Novamente de (2.1.5) e (2.1.6) obtém-se ainda,

$$\mathbb{E}\left(d_i^2 {\mathrm{e}'_i}^2\right) \leqslant \frac{1}{(K-1)^2}\sum_{j=1}^{K} j^2 p_j, \quad K \geqslant 2$$

donde,

$$\sum_{i=1}^{\infty}\mathbb{E}\left(d_i^2 {\mathrm{e}'_i}^2\right) \leqslant p_1\widetilde{N}(1)\sup_{i\geqslant 1}\left(d_i^2\right)+\sum_{K=2}^{\infty}\sum_{\left\{i:\ K-1<|d_i|^{-1}\leqslant K\right\}}(K-1)^{-2}\sum_{j=1}^{K} j^2 p_j$$

ou seja,

$$\sum_{i=1}^{\infty}\mathbb{E}\left(d_i^2 {\mathrm{e}'_i}^2\right) \leqslant p_1\widetilde{N}(1)\sup_{i\geqslant 1}\left(d_i^2\right)+\sum_{K=2}^{\infty}\Big(\widetilde{N}(K)-\widetilde{N}(K-1)\Big)(K-1)^{-2}\sum_{j=1}^{K} j^2 p_j.$$

Mas,

$$\sum_{K=2}^{\infty}\Big(\widetilde{N}(K)-\widetilde{N}(K-1)\Big)(K-1)^{-2}\sum_{j=1}^{K}j^2p_j=-p_1\widetilde{N}(1)+$$
$$+\sum_{K=2}^{\infty}p_1\widetilde{N}(K)\Big((K-1)^{-2}-K^{-2}\Big)-\sum_{j=2}^{\infty}\widetilde{N}(j-1)\left(\frac{j}{j-1}\right)^2p_j+$$
$$+\sum_{j=2}^{\infty}\left(\sum_{K=j}^{\infty}\widetilde{N}(K)\Big((K-1)^{-2}-K^{-2}\Big)\right)j^2p_j$$

e como $\widetilde{N}(K)=O\left(\frac{K^r}{\log^s K}\right)$ com $s>1$ tem-se

$$\widetilde{N}(K)\big((K-1)^{-2}-K^{-2}\big)=O\left(\frac{1}{K^{3-r}\log^s K}\right).$$

Então para $j\geqslant 2$ vem,

$$\sum_{K=j}^{\infty}\frac{1}{K^{3-r}\log^s K}\leqslant\frac{1}{\log^s 2}\sum_{K=j}^{\infty}\frac{1}{K^{3-r}}\leqslant\frac{1}{\log^s 2}\left(j^{r-3}+\int_j^{\infty}\frac{1}{t^{3-r}}\,dt\right)\leqslant$$
$$\leqslant\frac{1}{\log^s 2}\left(j^{r-2}+\frac{j^{r-2}}{2-r}\right)\leqslant\frac{j^{r-2}}{\log^s 2}\left(\frac{3-r}{2-r}\right)$$

ficando garantida a convergência da série $\sum_{K=2}^{\infty}\widetilde{N}(K)\Big((K-1)^{-2}-K^{-2}\Big)$. Por outro lado,

$$\sum_{j=2}^{\infty}\left(\sum_{K=j}^{\infty}\widetilde{N}(K)\Big((K-1)^{-2}-K^{-2}\Big)\right)j^2p_j\leqslant\sum_{j=2}^{\infty}c_4(r,s)j^rp_j$$

para alguma constante $c_4(r,s)>0$ pelo que as séries $\sum_{j=2}^{\infty}\widetilde{N}(j-1)\left(\frac{j}{j-1}\right)^2p_j$ e $\sum_{j=2}^{\infty}c_4(r,s)j^rp_j$ são convergentes por comparação com $\sum_{j=2}^{\infty}(j-1)^rp_j<\infty$. Portanto,

$$\sum_{i=1}^{\infty}\mathbb{E}\left(d_i^2\mathrm{e}_i'^2\right)<\infty$$

e a tese fica estabelecida pelo lema de Kronecker matricial com $\boldsymbol{A}_i=\boldsymbol{S}_i$ e $\boldsymbol{v}_i=\boldsymbol{d}_i\mathrm{e}_i$. ∎

Observação 2.1.1 Observe-se que a hipótese (b) do Teorema 2.1.3 pode ser refinada pondo para qualquer $s>1$,

$$N(K)=O\left(\frac{K^r}{\log K\log(\log K)\dots(\log_m K)^s}\right),\quad K\to\infty$$

onde $\log_m K=\log(\log(\dots\log(K)))$ para algum inteiro $m\geqslant 2$. Neste caso, o estudo das séries de Kolmogorov deverá ser efectuado a partir da ordem $\big[\exp_{(m-1)}1\big]+1$ onde $[x]$ indica o maior inteiro menor ou igual a x e $\exp_{(m-1)}1=\exp(\exp(\dots\exp(1)))$.

2.2 Condições de suficiência para erros estáveis

Iniciamos esta secção com um teorema que estende o resultado desenvolvido por Lai, Robbins e Wei (ver [53] e [54]) ao caso em que os erros têm valor médio não-nulo.

Teorema 2.2.1 *Se no modelo de regressão linear* (2.1.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $N(\eta, \nu^2)$ *com* $\eta \neq 0$.

(b) $\lim_{n\to\infty} \boldsymbol{S}_n^{-1} = \mathbf{O}$, $\limsup_{n\to\infty} \sum_{i=1}^{n} \left\| \boldsymbol{S}_n^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^T \right\| < \infty$ *e* $\left\| \sum_{i=1}^{\infty} \boldsymbol{S}_i^{-1} \boldsymbol{x}_i \right\| < \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Visto que $\mathrm{e}_i - \eta \sim N(0, \nu^2)$ a tese é uma consequência do resultado de [53] (ou [54]) e do lema de Kronecker matricial. ∎

Os próximos teoremas constituem uma alternativa aos resultados estabelecidos na anterior secção. Mais concretamente, vamos admitir unicamente que a sucessão de erros é i.i.d. α-estável ($\alpha < 2$) excluindo quaisquer hipóteses sobre os seus momentos absolutos.

Teorema 2.2.2 *Seja* $0 < \alpha < 2$, $\alpha \neq 1$ *e suponhamos que no modelo de regressão linear* (2.1.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_\alpha(\sigma, \lambda, \mu)$.

(b) $\lim_{n\to\infty} \boldsymbol{S}_n^{-1} = \mathbf{O}$, $\limsup_{n\to\infty} \sum_{i=1}^{n} \left\| \boldsymbol{S}_n^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^T \right\| < \infty$ *e* $\sum_{i=1}^{\infty} \left\| \boldsymbol{S}_i^{-1} \boldsymbol{x}_i \right\|^{\min(\alpha,1)} < \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. A identidade (2.1.4) garante que,

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \mathrm{e}_i = \boldsymbol{\beta} + \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \mathrm{e}_i^* + \mu\, \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i$$

com $\mathrm{e}_i^* \sim \mathcal{S}_\alpha(\sigma, \lambda, 0)$. Adoptando o mesmo raciocínio do Teorema 2.2.1 vem,

$$\mu\, \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \longrightarrow \mathbf{0}$$

quando $n \to \infty$ já que $\left\| \sum_{i=1}^{\infty} \boldsymbol{S}_i^{-1} \boldsymbol{x}_i \right\| < \infty$. Por outro lado, cada componente do vector $\sum_{i=1}^{n} \boldsymbol{S}_i^{-1} \boldsymbol{x}_i \mathrm{e}_i^*$ tem distribuição,

$$\mathcal{S}_\alpha \left(\sigma \left(\sum_{i=1}^{n} |d_{ij}|^\alpha \right)^{1/\alpha}, \frac{\lambda \sum_{i=1}^{n} \operatorname{sign}(d_{ij})\, |d_{ij}|^\alpha}{\sum_{i=1}^{n} |d_{ij}|^\alpha}, 0 \right), \quad j = 1, \ldots, \kappa$$

onde d_{ij} é a j-ésima componente de $\boldsymbol{S}_i^{-1}\boldsymbol{x}_i$. Visto que,

$$\sum_{i=1}^{\infty} |d_{ij}|^{\alpha} \leqslant \sum_{i=1}^{\infty} \left\| \boldsymbol{S}_i^{-1}\boldsymbol{x}_i \right\|^{\alpha}$$

temos a convergência quase certa de $\sum_{i=1}^{\infty} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i \mathrm{e}_i^*$ através do Teorema 1.4.5 e do lema de Kronecker matricial conclui-se que,

$$\boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \mathrm{e}_i^* \xrightarrow{\text{q.c.}} \mathbf{0}. \qquad \blacksquare$$

No que se segue, usaremos a notação $|\boldsymbol{x}|_{\min} = \min\big(|x_1|, \ldots, |x_\kappa|\big)$ onde $\boldsymbol{x}$ é um κ-vector não-estocástico.

Teorema 2.2.3 *Suponhamos que no modelo de regressão linear* (2.1.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_1(\sigma, \lambda, \mu)$.

(b) $\lim_{n\to\infty} \boldsymbol{S}_n^{-1} = \mathbf{O}$, $\limsup_{n\to\infty} \sum_{i=1}^{n} \left|\!\left|\!\left| \boldsymbol{S}_n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T \right|\!\right|\!\right| < \infty$ *e* $\sum_{i=1}^{\infty} \left\| \boldsymbol{S}_i^{-1}\boldsymbol{x}_i \right\| \log \left| \boldsymbol{S}_i^{-1}\boldsymbol{x}_i \right|_{\min}$ *converge.*

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Analogamente, da identidade (2.1.4) vem,

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \mathrm{e}_i = \boldsymbol{\beta} + \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \mathrm{e}_i^* + \mu\, \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i$$

com $\mathrm{e}_i^* \sim \mathcal{S}_1(\sigma, \lambda, 0)$. Visto que $\mu\, \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \longrightarrow \mathbf{0}$ quando $n \to \infty$ resta mostrar que $\left\| \sum_{i=1}^{\infty} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i \mathrm{e}_i^* \right\|$ $< \infty$. Cada componente do vector $\sum_{i=1}^{n} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i \mathrm{e}_i^*$ tem distribuição,

$$\mathcal{S}_1\left(\sigma \sum_{i=1}^{n} |d_{ij}|, \frac{\lambda \sum_{i=1}^{n} \operatorname{sign}(d_{ij})\, |d_{ij}|}{\sum_{i=1}^{n} |d_{ij}|}, -\frac{2\sigma\lambda}{\pi} \sum_{i=1}^{n} d_{ij} \log |d_{ij}| \right), \quad j = 1, \ldots, \kappa$$

onde d_{ij} é a j-ésima componente de $\boldsymbol{S}_i^{-1}\boldsymbol{x}_i$. Visto que $\lim_{i\to\infty} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i = \mathbf{0}$ (observe-se que por hipótese $\limsup_{i\to\infty} \left|\!\left|\!\left| \boldsymbol{S}_i^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T \right|\!\right|\!\right| < \infty$) então o Teorema 1.4.5 e o lema de Kronecker matricial permitem concluir que,

$$\boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \mathrm{e}_i^* \xrightarrow{\text{q.c.}} \mathbf{0}. \qquad \blacksquare$$

Observação 2.2.1 Se os erros $\mathrm{e}_1, \mathrm{e}_2, \ldots$ forem i.i.d. estritamente estáveis com índice de estabilidade $0 < \alpha < 2$ então a consistência forte do EMQ permanece verdadeira quando $\lim_{n\to\infty} \boldsymbol{S}_n^{-1} = \mathbf{O}$, $\limsup_{n\to\infty} \sum_{i=1}^{n} \left|\!\left|\!\left| \boldsymbol{S}_n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T \right|\!\right|\!\right| < \infty$ e $\sum_{i=1}^{\infty} \left\| \boldsymbol{S}_i^{-1}\boldsymbol{x}_i \right\|^{\alpha} < \infty$.

Os resultados apresentados anteriormente podem ser estendidos à situação em que as variáveis aleatórias $\mathrm{e}_1, \mathrm{e}_2, \ldots$ são (apenas) independentes e tais que $\mathrm{e}_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, \mu_i)$. Neste sentido, finalizamos este capítulo com esta generalização onde a hipótese de idêntica distribuição dos erros $\mathrm{e}_1, \mathrm{e}_2, \ldots$ é completamente excluída. Para uma melhor exposição das nossas ideias e dos resultados a elas associados comecemos por anunciar os casos $0 < \alpha \leqslant 2$, $\alpha \neq 1$.

Teorema 2.2.4 *Seja $0 < \alpha \leqslant 2$, $\alpha \neq 1$ e suponhamos que no modelo de regressão linear* (2.1.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes e tais que* $\mathrm{e}_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, \mu_i)$.

(b) $\lim_{n\to\infty} \boldsymbol{S}_n^{-1} = \mathbf{O}$, $\limsup_{n\to\infty} \sum_{i=1}^{n} \left\|\boldsymbol{S}_n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T\right\| < \infty$, $\sum_{i=1}^{\infty} \left(\sigma_i \left\|\boldsymbol{S}_i^{-1}\boldsymbol{x}_i\right\|\right)^\alpha < \infty$ *e* $\left\|\sum_{i=1}^{\infty} \mu_i \boldsymbol{S}_i^{-1}\boldsymbol{x}_i\right\| < \infty$.

então $\widetilde{\boldsymbol{\beta}}$ é fortemente consistente.

Demonstração. De (2.1.4) apenas temos que mostrar a convergência q.c. da série $\sum_{i=1}^{\infty} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i\mathrm{e}_i$ já que a versão matricial do Lema de Kronecker estabelece tudo o resto. Mas cada componente do vector $\sum_{i=1}^{n} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i\mathrm{e}_i$ tem distribuição,

$$\mathcal{S}_\alpha\left(\left(\sum_{i=1}^{n}\left(\sigma_i\left|d_{ij}\right|\right)^\alpha\right)^{1/\alpha}, \frac{\sum_{i=1}^{n}\left(\sigma_i\left|d_{ij}\right|\right)^\alpha\lambda_i\mathrm{sign}(d_{ij})}{\sum_{i=1}^{n}\left(\sigma_i\left|d_{ij}\right|\right)^\alpha}, \sum_{i=1}^{n}\mu_i d_{ij}\right), \quad j = 1, \ldots, \kappa$$

donde pelo Teorema 1.4.5 concluímos que $\sum_{i=1}^{\infty} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i\mathrm{e}_i$ converge quase certamente ficando, assim, a consistência forte do estimador estabelecida. ∎

Terminamos com a apresentação do caso $\alpha = 1$.

Teorema 2.2.5 *Suponhamos que no modelo de regressão linear* (2.1.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes e tais que* $\mathrm{e}_i \sim \mathcal{S}_1(\sigma_i, \lambda_i, \mu_i)$.

(b) $\lim_{n\to\infty} \boldsymbol{S}_n^{-1} = \mathbf{O}$, $\limsup_{n\to\infty} \sum_{i=1}^{n} \left\|\boldsymbol{S}_n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T\right\| < \infty$, $\sum_{i=1}^{\infty} \sigma_i \left\|\boldsymbol{S}_i^{-1}\boldsymbol{x}_i\right\| \log\left|\boldsymbol{S}_i^{-1}\boldsymbol{x}_i\right|_{\min}$ *converge e* $\left\|\sum_{i=1}^{\infty} \mu_i \boldsymbol{S}_i^{-1}\boldsymbol{x}_i\right\| < \infty$.

então $\widetilde{\boldsymbol{\beta}}$ é fortemente consistente.

Demonstração. Utilizando a identidade (2.1.4) será suficiente demonstrar a convergência quase certa da série,

$$\sum_{i=1}^{\infty} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i\mathrm{e}_i \tag{2.2.1}$$

pois o lema de Kronecker matricial garantirá a consistência forte do EMQ. Visto que cada componente do vector $\sum_{i=1}^{n} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i \mathrm{e}_i$ tem distribuição,

$$\mathcal{S}_1\left(\sum_{i=1}^{n}\sigma_i\left|d_{ij}\right|, \frac{\sum_{i=1}^{n}\sigma_i\left|d_{ij}\right|\lambda_i \mathrm{sign}(d_{ij})}{\sum_{i=1}^{n}\sigma_i\left|d_{ij}\right|}, \sum_{i=1}^{n}\left(\mu_i - \frac{2\sigma_i\lambda_i}{\pi}\log\left|d_{ij}\right|\right)d_{ij}\right), \quad j = 1,\ldots,\kappa$$

onde d_{ij} é a j-ésima componente de $\boldsymbol{S}_i^{-1}\boldsymbol{x}_i$. A convergência quase certa da série (2.2.1) resulta do Teorema 1.4.5 uma vez que $\lim\limits_{i\to\infty} \boldsymbol{S}_i^{-1}\boldsymbol{x}_i = \mathbf{0}$. ■

Observação 2.2.2 Se os erros $\mathrm{e}_1, \mathrm{e}_2, \ldots$ forem independentes, α-estáveis e simétricos (em relação à origem) então a consistência forte do EMQ é válida quando

$$\lim_{n\to\infty} \boldsymbol{S}_n^{-1} = \mathbf{O}, \quad \limsup_{n\to\infty}\sum_{i=1}^{n}\left\|\boldsymbol{S}_n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T\right\| < \infty \quad \text{e} \quad \sum_{i=1}^{\infty}\left(\sigma_i\left\|\boldsymbol{S}_i^{-1}\boldsymbol{x}_i\right\|\right)^{\alpha} < \infty.$$

Capítulo 3

Na estabilidade dos erros

No que se vai seguir e em capítulos posteriores, assumiremos sempre que a matriz de modelo $\boldsymbol{X}_n = \left[x_{ij}\right]_{\substack{i=1,\ldots,n \\ j=1,\ldots,\kappa}}$ tem característica κ e excluiremos também qualquer hipótese de multicolinearidade quase-exacta da matriz $\boldsymbol{X}_n$ (conf. Secção 1.6). Até ao final deste texto, destacaremos ainda o índice n do número de linhas da matriz de modelo e do número de componentes do vector dos erros uma vez que, em resultados futuros, vão acontecer passagens ao limite em n.

Neste capítulo, supondo que os erros são α-estáveis, temos como principal objectivo estabelecer condições sobre o raio espectral da matriz $\left(\boldsymbol{X}_n^T\boldsymbol{X}_n\right)^{-1}$ para que o EMQ do vector parâmetro $\boldsymbol{\beta}$ referente ao modelo de regressão linear,

$$\mathbf{Y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \mathbf{e}_n \tag{3.0.1}$$

seja fortemente consistente. Para alcançarmos a consistência forte do EMQ, usaremos uma importante estimativa para a diferença entre o EMQ e o vector parâmetro $\boldsymbol{\beta}$, obtida através da Proposição 1.2.3. Com efeito, efectuando a substituição do vector das observações $\mathbf{Y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \mathbf{e}_n$ em (1.6.9) vem,

$$\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\boldsymbol{X}_n^T\boldsymbol{X}_n\right)^{-1}\boldsymbol{X}_n^T\mathbf{e}_n$$

donde aplicando a Proposição 1.2.3 com $\boldsymbol{A} = \boldsymbol{X}_n$, $\mathbb{X}_n = \mathrm{Im}(\boldsymbol{X}_n)$ e $\boldsymbol{x} = \mathbf{e}_n$ tem-se automaticamente,

$$\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2 \leqslant \rho\Big((\boldsymbol{X}_n^T\boldsymbol{X}_n)^{-1}\Big)\,\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2\,. \tag{3.0.2}$$

Sinteticamente, representaremos o raio espectral $\rho\Big((\boldsymbol{X}_n^T\boldsymbol{X}_n)^{-1}\Big)$ simplesmente por ρ_n.

O uso prático de modelos lineares de efeitos fixos tais como os estudados aqui é bastante comum. Uma situação típica ocorre em designs de modelos lineares múltiplos. Veja-se [61], onde em cada tratamento de uma base de design temos regressão nalgumas variáveis. Podemos considerar que, para cada tratamento, são feitas observações igualmente espaçadas nas quais as variáveis de controlo tomam como valores potências sucessivas de i para $i = 1, 2, \ldots, n$. Exemplos desta metodologia podem ser vistos ainda em [69] e [80].

3.1 O caso normal

Na secção 1.4 do primeiro capítulo vimos (conf. Exemplo 1.4.1) que qualquer v.a. com distribuição $N(\eta, \nu^2)$ é estável, ou seja, mais concretamente,

$$\mathcal{S}_2(\sigma, \lambda, \mu) = N(\mu, 2\sigma^2) \qquad (\sigma > 0)$$

onde o parâmetro de distorção λ é totalmente irrelevante. Nesta secção incidiremos os holofotes sobre este caso especial das distribuições estáveis conseguindo assim e numa primeira instância, um resultado igual ao obtido por Lai, Robbins e Wei (ver [53] e [54]). Apresentaremos, em seguida, a extensão natural à situação em que o valor médio dos erros é não-nulo e desenvolveremos também os correspondentes resultados de consistência em média quadrática.

Na Observação 1.4.4 referimos que as funções de distribuição de v.a.'s α-estáveis não-degeneradas são absolutamente contínuas e as suas densidades infinitamente diferenciáveis em todo o ponto da recta real. Mas excluindo alguns casos especiais (de entre eles o gaussiano) a representação das densidades de uma v.a. estável apenas pode ser feita através de complicadas funções especiais. Contudo, existem desenvolvimentos assimptóticos de densidades α-estáveis numa vizinhança da origem ou de infinito (ver [41] ou [93] para mais detalhes).

Teorema 3.1.1 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_2(\sigma, \lambda, 0)$ *com* $\sigma > 0$.

(b) $\rho_n = o(1)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Aplicando a Proposição 1.2.1 a $\boldsymbol{A} = \boldsymbol{X}_n$ e $\boldsymbol{x} = \mathbf{e}_n$ existirá uma base ortonormada

$$\begin{gathered}\boldsymbol{w}_1(n) = \big(w_{11}(n), w_{21}(n), \ldots, w_{n1}(n)\big)\\ \vdots \\ \boldsymbol{w}_\kappa(n) = \big(w_{1\kappa}(n), w_{2\kappa}(n), \ldots, w_{n\kappa}(n)\big)\end{gathered}$$

tal que a v.a. $\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2$ é expressa por,

$$\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2 = \Big(\mathrm{S}_{n1}(n)\Big)^2 + \ldots + \Big(\mathrm{S}_{n\kappa}(n)\Big)^2 \tag{3.1.1}$$

onde a sucessão $\big\{\mathrm{S}_{ni}(n),\ n \geqslant 1\big\}$ está definida pelo arranjo triangular de v.a.'s,

$$\begin{aligned}\mathrm{S}_{1i}(1) &= w_{1i}(1)\mathrm{e}_1\\ \mathrm{S}_{2i}(2) &= w_{1i}(2)\mathrm{e}_1 + w_{2i}(2)\mathrm{e}_2\\ &\ \ \vdots\\ \mathrm{S}_{ni}(n) &= w_{1i}(n)\mathrm{e}_1 + w_{2i}(n)\mathrm{e}_2 + \ldots + w_{ni}(n)\mathrm{e}_n\end{aligned} \tag{3.1.2}$$

para todo o $i = 1, \ldots, \kappa$. Utilizando a desigualdade de Levy (ver Teorema 1.1.6) a cada termo da sucessão $\{\mathrm{S}_{ni}(n),\, n \geqslant 1\}$ obtém-se,

$$\begin{aligned}
\mathbb{P}\left\{\max_{1\leqslant j\leqslant 1} |\mathrm{S}_{ji}(1)| \geqslant \varepsilon\right\} &\leqslant 2\cdot\mathbb{P}\Big\{\,|\mathrm{S}_{1i}(1)| \geqslant \varepsilon\Big\} \leqslant 2\cdot\frac{\mathbb{E}\big(\,|\mathrm{S}_{1i}(1)|\,\big)}{\varepsilon} \leqslant \\
&\leqslant 2\cdot\frac{\mathbb{E}\big(\,\|\boldsymbol{P}_{\mathbb{X}_1}\mathbf{e}_1\|\,\big)}{\varepsilon} = \frac{4\sigma\,\Gamma\left(\frac{\kappa+1}{2}\right)}{\varepsilon\,\Gamma\left(\frac{\kappa}{2}\right)} \\
\mathbb{P}\left\{\max_{1\leqslant j\leqslant 2} |\mathrm{S}_{ji}(2)| \geqslant \varepsilon\right\} &\leqslant 2\cdot\mathbb{P}\Big\{\,|\mathrm{S}_{2i}(2)| \geqslant \varepsilon\Big\} \leqslant 2\cdot\frac{\mathbb{E}\big(\,|\mathrm{S}_{2i}(2)|\,\big)}{\varepsilon} \leqslant \\
&\leqslant 2\cdot\frac{\mathbb{E}\big(\,\|\boldsymbol{P}_{\mathbb{X}_2}\mathbf{e}_2\|\,\big)}{\varepsilon} = \frac{4\sigma\,\Gamma\left(\frac{\kappa+1}{2}\right)}{\varepsilon\,\Gamma\left(\frac{\kappa}{2}\right)} \\
&\vdots \\
\mathbb{P}\left\{\max_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)| \geqslant \varepsilon\right\} &\leqslant 2\cdot\mathbb{P}\Big\{\,|\mathrm{S}_{ni}(n)| \geqslant \varepsilon\Big\} \leqslant 2\cdot\frac{\mathbb{E}\big(\,|\mathrm{S}_{ni}(n)|\,\big)}{\varepsilon} \leqslant \\
&\leqslant 2\cdot\frac{\mathbb{E}\big(\,\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|\,\big)}{\varepsilon} = \frac{4\sigma\,\Gamma\left(\frac{\kappa+1}{2}\right)}{\varepsilon\,\Gamma\left(\frac{\kappa}{2}\right)}
\end{aligned}$$

para qualquer $i = 1, \ldots, \kappa$ pois o teorema de Cochran (ver Teorema 1.2.10) garante que $\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2$ $\sim 2\sigma^2\chi^2(\kappa)$, $\forall n \in \mathbb{N}$. Então,

$$\mathbb{P}\left\{\max_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)| \geqslant \varepsilon\right\} \leqslant \frac{4\sigma\,\Gamma\left(\frac{\kappa+1}{2}\right)}{\varepsilon\,\Gamma\left(\frac{\kappa}{2}\right)}, \quad \forall n \in \mathbb{N}$$

e escolhendo a subsucessão $\max\limits_{1\leqslant j\leqslant \xi_n} |\mathrm{S}_{ji}(\xi_n)|$ de $\max\limits_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)|$ que nos dá,

$$\lim_{n\to\infty}\left(\max_{1\leqslant j\leqslant \xi_n} |\mathrm{S}_{ji}(\xi_n)|\right) = \limsup_{n\to\infty}\left(\max_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)|\right)$$

vem,

$$\mathbb{P}\left\{\max_{1\leqslant j\leqslant \xi_n} |\mathrm{S}_{ji}(\xi_n)| \geqslant \varepsilon\right\} \leqslant \frac{4\sigma\,\Gamma\left(\frac{\kappa+1}{2}\right)}{\varepsilon\,\Gamma\left(\frac{\kappa}{2}\right)}, \quad \forall n \in \mathbb{N}.$$

Efectuando a passagem ao limite em n surge,

$$\mathbb{P}\left\{\limsup_{n\to\infty}\left(\max_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)|\right) \geqslant \varepsilon\right\} \leqslant \frac{4\sigma\,\Gamma\left(\frac{\kappa+1}{2}\right)}{\varepsilon\,\Gamma\left(\frac{\kappa}{2}\right)}$$

e fazendo $\varepsilon \to \infty$ concluí-se que $\limsup\limits_{n\to\infty}\left(\max\limits_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)|\right)$ existe finito quase certamente para todo o $i = 1, \ldots, \kappa$. Consequentemente,

$$\limsup_{n\to\infty}\ \|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2$$

existe finito quase certamente.[1] A estimativa (3.0.2) e a hipótese sobre o raio espectral estabelecem a tese. ∎

[1]Observemos que $\limsup\limits_{n\to\infty}\ \|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2 > 0$ q.c. uma vez que $\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2 \sim 2\sigma^2\chi^2(\kappa)$, $\forall n \in \mathbb{N}$.

A consistência em média quadrática está também assegurada sob idêntica hipótese para o raio espectral de $\left(\boldsymbol{X}_n^T\boldsymbol{X}_n\right)^{-1}$.

Teorema 3.1.2 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_2(\sigma, \lambda, 0)$ *com* $\sigma > 0$.

(b) $\rho_n = o(1)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. A estimativa (3.0.2) e o teorema de Cochran (conf. Teorema 1.2.10) asseguram que,

$$\mathbb{E}\left[\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2\right] \leqslant \rho_n\, \mathbb{E}\left[\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2\right] = 2\kappa\, \sigma^2\, \rho_n$$

o que conclui a prova já que por hipótese $\rho_n = o(1)$, $n \to \infty$. ∎

Observação 3.1.1 Note-se que, sob as hipóteses dos anteriores Teorema 3.1.1 e Teorema 3.1.2, as v.a.'s $\mathrm{e}_1, \mathrm{e}_2, \ldots$ são i.i.d. estritamente estáveis com expoente característico $\alpha = 2$.

Para situações em que as v.a.'s $\mathrm{e}_1, \mathrm{e}_2, \ldots$ não são estritamente estáveis consideremos o vector real $\boldsymbol{\mu} = (\mu, \ldots, \mu) \in \mathbb{R}^n$ bem como a sua projecção ortogonal $\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}$ sobre o espaço imagem $\mathbb{X}_n = \mathrm{Im}(\boldsymbol{X}_n)$. De seguida, apresentamos resultados de consistência forte e de consistência em média quadrática para os casos de estabilidade não-estrita.

Corolário 3.1.1 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_2(\sigma, \lambda, \mu)$ *com* $\sigma > 0$ *e* $\mu \neq 0$.

(b) ρ_n *é uma sucessão evanescente tal que* $\rho_n = o\left(\|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}\|^{-2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. O vector dos erros $\mathbf{e}_n$ pode ser decomposto em $\mathbf{e}_n = \boldsymbol{\mu} + \mathbf{e}_n^*$ onde as componentes do vector aleatório $\mathbf{e}_n^*$ (conhecido como *ruído branco* em teoria do Sinal) têm distribuição $N(0, 2\sigma^2)$. Visto que a projecção ortogonal $P_{\mathbb{X}_n}$ é linear obtém-se,

$$\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n = \boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu} + \boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n^*$$

e a desigualdade triangular garante que,

$$\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2 \leqslant 2\left(\|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}\|^2 + \|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n^*\|^2\right).$$

Então, a estimativa (3.0.2) pode ser reescrita na forma,

$$\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2 \leqslant 2\rho_n\, \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}\|^2 + 2\rho_n\, \|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n^*\|^2 \tag{3.1.3}$$

e do Teorema 3.1.1 tem-se $\rho_n\, \|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n^*\|^2 \xrightarrow{\text{q.c.}} 0$ quando $n \to \infty$. A tese fica estabelecida através da hipótese $\rho_n\, \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}\|^2 \longrightarrow 0$ quando $n \to \infty$. ∎

Corolário 3.1.2 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_2(\sigma, \lambda, \mu)$ *com* $\sigma > 0$ *e* $\mu \neq 0$.

(b) ρ_n *é uma sucessão evanescente tal que* $\rho_n = o\left(\|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}\|^{-2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. A tese é uma simples combinação da estimativa (3.1.3) e do Teorema 3.1.2. ∎

3.2 O caso $0 < \alpha < 2$

Em sintonia com o que foi previamente feito, assumiremos ao longo deste capítulo que o parâmetro de escala σ é positivo. No entanto, sempre que $\mathrm{Z} \sim \mathcal{S}_\alpha(0, \lambda, \mu)$ com $0 < \alpha \leqslant 2$ então $\mathrm{Z} = \mu$ (ver função característica (1.4.4)) donde todas as distribuições $\mathcal{S}_\alpha(0, \lambda, \mu)$ com $0 < \alpha \leqslant 2$ são degeneradas no ponto μ. Nesta situação, observe-se ainda que o parâmetro de distorção λ é completamente irrelevante. Deste modo, se $\mathrm{e}_1, \mathrm{e}_2, \ldots$ forem i.i.d. $\mathcal{S}_\alpha(0, \lambda, \mu)$ com $0 < \alpha \leqslant 2$, $\mu \neq 0$ então o EMQ $\widetilde{\boldsymbol{\beta}}$ é fortemente consistente e consistente em média quadrática para qualquer raio espectral ρ_n que verifique $\rho_n = o\left(\|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}\|^{-2}\right)$, $n \to \infty$. Evidentemente, se $\mu = 0$ então o EMQ é fortemente consistente e consistente em média quadrática para qualquer raio espectral ρ_n. Excluiremos situações como esta por não terem grande interesse na prática.

Teorema 3.2.1 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_\alpha(\sigma, \lambda, 0)$ *com* $0 < \alpha < 2$, $\alpha \neq 1$.

(b) $\rho_n = o\left(n^{1-\frac{2}{\alpha}}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Da Proposição 1.2.1 sabemos que existe uma base ortonormada $\{\boldsymbol{w}_1(n), \ldots, \boldsymbol{w}_n(n)\}$ de $\mathbb{R}^n$ tal que,

$$\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2 = \Big(\mathrm{S}_{n1}(n)\Big)^2 + \ldots + \Big(\mathrm{S}_{n\kappa}(n)\Big)^2$$

onde cada termo de $\big\{\mathrm{S}_{ni}(n),\, n \geqslant 1\big\}$ está definido por (3.1.2) para todo o $i = 1, \ldots, \kappa$. As propriedades das distribuições estáveis (conf. Proposição 1.4.1 e Proposição 1.4.3) asseguram que,

$$\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}} \sim \mathcal{S}_\alpha\Big(\bar{\sigma}(\alpha, n, p), \bar{\lambda}(\alpha, n, p), 0\Big), \quad p = 1, \ldots, n \tag{3.2.1}$$

onde,

$$\bar{\sigma}(\alpha, n, p) = \sigma\left(\sum_{j=p}^{n} \frac{|w_{ji}(n)|^\alpha}{n^{1-\frac{\alpha}{2}}}\right)^{\frac{1}{\alpha}} \quad \text{e} \quad \bar{\lambda}(\alpha, n, p) = \frac{\lambda \displaystyle\sum_{j=p}^{n} \operatorname{sign}\big(w_{ji}(n)\big)\,|w_{ji}(n)|^\alpha}{\displaystyle\sum_{j=p}^{n} |w_{ji}(n)|^\alpha}. \tag{3.2.2}$$

Escolhendo $r < \alpha$ vem,

$$\mathbb{P}\left\{\left|\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \delta\right\} \leqslant \frac{\mathbb{E}\left|\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}}\right|^r}{\delta^r} \leqslant \frac{\Big(\bar{\sigma}(\alpha,n,p)\cdot c_1\big(r,\bar{\lambda}(\alpha,n,p)\big)\Big)^r}{\delta^r}$$

para alguma constante $c_1\big(r,\bar{\lambda}(\alpha,n,p)\big) > 0$ dada por,

$$\Big(c_1\big(r,\bar{\lambda}(\alpha,n,p)\big)\Big)^r = \frac{2^{r-1}\Gamma\left(1-\frac{r}{\alpha}\right)}{r\int_0^\infty t^{-r-1}\sin^2 t\,dt}\left[1+\big(\bar{\lambda}(\alpha,n,p)\big)^2\tan^2\left(\frac{\alpha\pi}{2}\right)\right]^{\frac{r}{2\alpha}}\cdot$$
$$\cdot\cos\left(\frac{r}{\alpha}\arctan\left(\bar{\lambda}(\alpha,n,p)\tan\left(\frac{\alpha\pi}{2}\right)\right)\right)$$

(conf. Corolário 1.4.1). A Proposição 1.3.4 garante que $\sum_{j=p}^{n}\frac{|w_{ji}(n)|^\alpha}{n^{1-\frac{\alpha}{2}}} \leqslant 1$ o que implica $\bar{\sigma}(\alpha,n,p) \leqslant \sigma$ e,

$$\mathbb{P}\left\{\left|\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \delta\right\} \leqslant \frac{\Big(\bar{\sigma}(\alpha,n,p)\cdot c_1\big(r,\bar{\lambda}(\alpha,n,p)\big)\Big)^r}{\delta^r} \leqslant \frac{\Big(c_2(\alpha,\sigma,r)\Big)^r}{\delta^r} < 1$$

fixando $\delta > c_2(\alpha,\sigma,r)$ com,

$$\Big(c_2(\alpha,\sigma,r)\Big)^r = \frac{\sigma\, 2^{r-1}\Gamma\left(1-\frac{r}{\alpha}\right)}{r\int_0^\infty t^{-r-1}\sin^2 t\,dt}\cdot\left[1+\tan^2\left(\frac{\alpha\pi}{2}\right)\right]^{\frac{r}{2\alpha}}.$$

Pelo Teorema 1.1.7 tem-se para todo o $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\left\{\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \delta+\varepsilon\right\} &\leqslant \frac{1}{1-\frac{c_2(\alpha,\sigma,r)}{\delta^r}}\,\mathbb{P}\left\{\left|\sum_{j=1}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \varepsilon\right\}\\ &\leqslant \frac{1}{1-\frac{c_2(\alpha,\sigma,r)}{\delta^r}}\cdot\frac{\mathbb{E}\left|\sum_{j=1}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}}\right|^r}{\varepsilon^r}\\ &\leqslant \frac{\left(\frac{c_2(\alpha,\sigma,r)}{\delta}\right)^r}{1-\left(\frac{c_2(\alpha,\sigma,r)}{\delta}\right)^r}\cdot\left(\frac{\delta}{\varepsilon}\right)^r.\end{aligned}$$

Naturalmente, podemos efectuar os cálculos anteriores para cada termo,

$$\sum_{j=p}^{m} \frac{w_{ji}(m)\mathrm{e}_j}{m^{\frac{1}{\alpha}-\frac{1}{2}}},\quad p=1,\ldots,m$$

com $m = 1,\ldots,n-1$ e concluir que,

$$\mathbb{P}\left\{\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \delta+\varepsilon\right\} \leqslant \frac{\left(\frac{c_2(\alpha,\sigma,r)}{\delta}\right)^r}{1-\left(\frac{c_2(\alpha,\sigma,r)}{\delta}\right)^r}\cdot\left(\frac{\delta}{\varepsilon}\right)^r,\quad \forall n\in\mathbb{N}.$$

A tese fica estabelecida raciocinando como na parte final da demonstração do Teorema 3.1.1. ■

A consistência em momento de ordem r do EMQ é analisada de seguida.

Teorema 3.2.2 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_\alpha(\sigma, \lambda, 0)$ *com* $0 < \alpha < 2$, $\alpha \neq 1$.

(b) $\rho_n = o\left(n^{1-\frac{2}{\alpha}}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em momento de ordem* r *com* $r < \alpha$.

Demonstração. A estimativa (3.0.2) pode ser reescrita como (conf. Proposição 1.3.3),

$$\mathbb{E}\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^r\right) \leqslant (\rho_n)^{\frac{r}{2}}\, \mathbb{E}\Big(|\mathrm{S}_{n1}(n)|^r + \ldots + |\mathrm{S}_{n\kappa}(n)|^r \Big) \tag{3.2.3}$$

onde cada termo da sucessão $\left\{\mathrm{S}_{ni}(n),\ n \geqslant 1\right\}$ está definido por (3.1.2) para todo o $i = 1, \ldots, \kappa$. Visto que para todo o $n \in \mathbb{N}$,

$$\sum_{j=1}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}} \sim \mathcal{S}_\alpha\Big(\bar{\sigma}(\alpha, n, 1), \bar{\lambda}(\alpha, n, 1), 0\Big)$$

em que $\bar{\sigma}(\alpha, n, 1)$ e $\bar{\lambda}(\alpha, n, 1)$ são dados por (3.2.2) tem-se para cada $i = 1, \ldots, \kappa$

$$\mathbb{E}\left|\sum_{j=1}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{n^{\frac{1}{\alpha}-\frac{1}{2}}}\right|^r \leqslant \Big(c_2(\alpha, \sigma, r)\Big)^r, \quad \forall n \in \mathbb{N}.$$

A conclusão segue da hipótese sobre o raio espectral ρ_n. ∎

Corolário 3.2.1 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_\alpha(\sigma, \lambda, \mu)$ *com* $0 < \alpha < 2$, $\alpha \neq 1$.

(b) $\rho_n = o\left(\dfrac{1}{n^{\frac{2}{\alpha}-1} + \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}\|^2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. A demonstração é uma consequência do Teorema 3.2.1 e da estimativa (3.1.3). ∎

Corolário 3.2.2 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_\alpha(\sigma, \lambda, \mu)$ *com* $0 < \alpha < 2$, $\alpha \neq 1$.

(b) $\rho_n = o\left(\dfrac{1}{n^{\frac{2}{\alpha}-1} + \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}\|^2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em momento de ordem* r *com* $r < \alpha$.

Demonstração. A tese resulta da combinação do Teorema 3.2.2 e da estimativa (3.1.3). ∎

Terminamos esta secção com o último caso $\alpha = 1$. Quando $\mathrm{e}_1, \mathrm{e}_2, \ldots$ são i.i.d. $\mathcal{S}_1(\sigma, 0, 0)$ os cálculos utilizados no Teorema 3.2.1 e no Teorema 3.2.2 permanecem válidos passo a passo (conf. Corolário 1.4.1). Além disso, a estimativa (3.1.3) ainda é verdadeira o que estabelece os,

Teorema 3.2.3 *Se no modelo de regressão linear* (3.0.1),

(a) $e_1, e_2, \ldots$ *são i.i.d.* $\mathcal{S}_1(\sigma, 0, \mu)$.

(b) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Teorema 3.2.4 *Se no modelo de regressão linear* (3.0.1),

(a) $e_1, e_2, \ldots$ *são i.i.d.* $\mathcal{S}_1(\sigma, 0, \mu)$.

(b) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em momento de ordem* r *com* $r < 1$.

Pelo Teorema 1.4.3 as probabilidades de cauda de uma v.a. $Z \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$ com $0 < \alpha < 2$ têm o seguinte comportamento assimptótico,

$$\lim_{z \to \infty} z^\alpha \mathbb{P}\big(Z > z\big) = c(\alpha) \frac{1+\lambda}{2} \, \sigma^\alpha \tag{3.2.4}$$

e

$$\lim_{z \to \infty} z^\alpha \mathbb{P}\big(Z < -z\big) = c(\alpha) \frac{1-\lambda}{2} \, \sigma^\alpha \tag{3.2.5}$$

com $c(\alpha) = \left(\int_0^\infty t^{-\alpha} \sin t \, dt\right)^{-1}$. Visto que a função de distribuição de $|Z|$ é dada por,

$$F_{|Z|}(z) = F_Z(z) - F_Z(-z), \qquad z \geqslant 0$$

e $F_{|Z|}(z) = 0$ se $z < 0$ obtém-se de (3.2.4) e (3.2.5),

$$\overline{F}_{|Z|}(z) = \overline{F}_Z(z) + F_Z(-z) \sim c(\alpha) \, \sigma^\alpha \, z^{-\alpha}, \quad z \to \infty.$$

Este comportamento de cauda conduz-nos a $\mathbb{E}\,|Z|^r < \infty$ para $0 < r < \alpha$ porque $\mathbb{E}\,|Z|^r = \int_0^\infty \mathbb{P}\big\{ |Z|^r > z\big\} \, dz$ e assim podemos completar o estudo dos restantes casos.

Observação 3.2.1 Dada uma v.a. $Z \sim \mathcal{S}_\alpha(\sigma, \lambda, \mu)$ com $0 < \alpha < 2$ obtém-se das expressões (3.2.4) e (3.2.5),

$$\overline{F}_{Z^2}(z) = \overline{F}_Z\left(\sqrt{z}\right) + F_Z\left(-\sqrt{z}\right) \sim c(\alpha) \, \sigma^\alpha \, z^{-\frac{\alpha}{2}}, \quad z \to \infty.$$

Então, podemos afirmar que a distribuição de Z^2 tem cauda equivalente a uma distribuição estável com expoente característico $\frac{\alpha}{2}$. Concretamente, a função F_{Z^2} (f.d. de Z^2) e $\mathcal{S}_{\frac{\alpha}{2}}(\sigma', 1, 0)$ onde,

$$\sigma' = \begin{cases} 4\sigma^2 \left[\dfrac{\sqrt{\pi} \, \cos\left(\frac{\alpha\pi}{4}\right)}{\Gamma\left(\frac{1-\alpha}{2}\right) \, \cos\left(\frac{\alpha\pi}{2}\right)} \right]^{\frac{2}{\alpha}} & \text{se} \quad \alpha \neq 1 \\[2em] \dfrac{2\sigma^2}{\pi} & \text{se} \quad \alpha = 1 \end{cases}$$

têm comportamento assimptótico de cauda equivalente.

Teorema 3.2.5 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.* $\mathcal{S}_1(\sigma, \lambda, \mu)$ *com* $\lambda \neq 0$.

(b) $\rho_n = o\left(\dfrac{1}{n \log^2 n}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Das Proposições 1.4.1 e 1.4.3 obtém-se para todo $i = 1, \ldots, \kappa$,

$$\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\sqrt{n}\log n} \sim \mathcal{S}_1\Big(\bar{\sigma}(n,p), \bar{\lambda}(n,p), \bar{\mu}(n,p)\Big), \quad p = 1, \ldots, n$$

em que,

$$\bar{\sigma}(n,p) = \sigma \sum_{j=p}^{n} \frac{|w_{ji}(n)|}{\sqrt{n}\log n}, \quad \bar{\lambda}(n,p) = \frac{\lambda \displaystyle\sum_{j=p}^{n} \operatorname{sign}\big(w_{ji}(n)\big)\,|w_{ji}(n)|}{\displaystyle\sum_{j=p}^{n} |w_{ji}(n)|} \tag{3.2.6}$$

e

$$\begin{aligned} \bar{\mu}(n,p) = \frac{w_{pi}(n)}{\sqrt{n}\log n} + \ldots + \frac{w_{ni}(n)}{\sqrt{n}\log n} - \\ - \frac{2\sigma\lambda}{\pi}\left(\frac{w_{pi}(n)\log|w_{pi}(n)|}{\sqrt{n}\log n} + \ldots + \frac{w_{ni}(n)\log|w_{ni}(n)|}{\sqrt{n}\log n}\right). \end{aligned} \tag{3.2.7}$$

Pela Proposição 1.3.5 temos que $\bar{\sigma}(n,p) \leqslant \sigma$, $\left|\bar{\lambda}(n,p)\right| \leqslant |\lambda|$, $\left|\bar{\lambda}(n,p)\right| \leqslant |\lambda|$ e $|\bar{\mu}(n,p)| \leqslant \frac{1}{\log 2} + \frac{\sigma|\lambda|}{\pi}$. Visto que,

$$\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\sqrt{n}\log n} \stackrel{\mathrm{d}}{=} \bar{\sigma}(n,p)\mathrm{Z}_n + \frac{2}{\pi}\bar{\lambda}(n,p)\bar{\sigma}(n,p)\log\big(\bar{\sigma}(n,p)\big) + \bar{\mu}(n,p)$$

com,

$$\begin{aligned} \mathrm{Z}_n \stackrel{\mathrm{d}}{=} \left(\frac{1+\bar{\lambda}(n,p)}{2}\right)\mathrm{A}_n - \left(\frac{1-\bar{\lambda}(n,p)}{2}\right)\mathrm{B}_n + \\ + \left(\frac{1+\bar{\lambda}(n,p)}{\pi}\right)\log\left(\frac{1+\bar{\lambda}(n,p)}{2}\right) - \left(\frac{1-\bar{\lambda}(n,p)}{\pi}\right)\log\left(\frac{1-\bar{\lambda}(n,p)}{2}\right) \end{aligned}$$

onde $\{\mathrm{A}_n,\, n \geqslant 1\}$ e $\{\mathrm{B}_n,\, n \geqslant 1\}$ denotam duas sucessões de v.a.'s tais que A_n e B_n são independentes com distribuição comum $\mathcal{S}_\alpha(1,1,0)$ para todo o $n \in \mathbb{N}$, tem-se para cada $i = 1, \ldots, \kappa$ e para qualquer $0 < r < 1$,

$$\begin{aligned} \mathbb{E}\left|\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\sqrt{n}\log n}\right|^r &= \mathbb{E}\left|\bar{\sigma}(n,p)\mathrm{Z}_n + \frac{2}{\pi}\bar{\lambda}(n,p)\bar{\sigma}(n,p)\log\big(\bar{\sigma}(n,p)\big) + \bar{\mu}(n,p)\right|^r \\ &\leqslant \mathbb{E}\left(\bar{\sigma}(n,p)\,|\mathrm{Z}_n| + \frac{2}{\pi}\left|\bar{\lambda}(n,p)\right|\left|\bar{\sigma}(n,p)\log\big(\bar{\sigma}(n,p)\big)\right| + |\bar{\mu}(n,p)|\right)^r \\ &\leqslant \mathbb{E}\left(\sigma\,|\mathrm{Z}_n| + \frac{2\,|\lambda|\,\sigma_{\max}}{\pi} + \frac{1}{\log 2} + \frac{\sigma\,|\lambda|}{\pi}\right)^r \\ &\leqslant \mathbb{E}\left[\sigma^r\,|\mathrm{Z}_n|^r + \left(\frac{2\,|\lambda|\,\sigma_{\max}}{\pi}\right)^r + \left(\frac{1}{\log 2}\right)^r + \left(\frac{\sigma\,|\lambda|}{\pi}\right)^r\right] \\ &= \sigma^r\mathbb{E}\big(\,|\mathrm{Z}_n|^r\,\big) + \left(\frac{2\,|\lambda|\,\sigma_{\max}}{\pi}\right)^r + \left(\frac{1}{\log 2}\right)^r + \left(\frac{\sigma\,|\lambda|}{\pi}\right)^r, \quad p = 1, \ldots, n \end{aligned}$$

com $\sigma_{\max} = \max\left(e^{-1}, \sigma \log \sigma\right)$. Analogamente,

$$\begin{aligned}
\mathbb{E}\left(|\mathrm{Z}_n|^r \right) &= \mathbb{E}\Bigg| \left(\frac{1+\bar{\lambda}(n,p)}{2} \right) \mathrm{A}_n - \left(\frac{1-\bar{\lambda}(n,p)}{2} \right) \mathrm{B}_n + \\
&\quad + \left(\frac{1+\bar{\lambda}(n,p)}{\pi} \right) \log \left(\frac{1+\bar{\lambda}(n,p)}{2} \right) - \left(\frac{1-\bar{\lambda}(n,p)}{\pi} \right) \log \left(\frac{1-\bar{\lambda}(n,p)}{2} \right) \Bigg|^r \\
&\leqslant \mathbb{E}\Bigg[\frac{1+\left|\bar{\lambda}(n,p)\right|}{2} \, |\mathrm{A}_n| + \frac{1+\left|\bar{\lambda}(n,p)\right|}{2} \, |\mathrm{B}_n| + \\
&\quad + \frac{2}{\pi} \left| \frac{1+\bar{\lambda}(n,p)}{2} \cdot \log \left(\frac{1+\bar{\lambda}(n,p)}{2} \right) \right| + \frac{2}{\pi} \left| \frac{1-\bar{\lambda}(n,p)}{\pi} \log \left(\frac{1-\bar{\lambda}(n,p)}{2} \right) \right| \Bigg]^r \\
&\leqslant \mathbb{E} \left(\frac{1+|\lambda|}{2} \, |\mathrm{A}_n| + \frac{1+|\lambda|}{2} \, |\mathrm{B}_n| + \frac{4}{e\pi} \right)^r \\
&\leqslant \left(\frac{1+|\lambda|}{2} \right)^r \mathbb{E}\left(|\mathrm{A}_n|^r \right) + \left(\frac{1+|\lambda|}{2} \right)^r \mathbb{E}\left(|\mathrm{B}_n|^r \right) + \left(\frac{4}{e\pi} \right)^r \\
&\leqslant 2c_3 \left(\frac{1+|\lambda|}{2} \right)^r + \left(\frac{4}{e\pi} \right)^r
\end{aligned}$$

para alguma constante $c_3 > 0$ independente de n uma vez que a distribuição de A_n e B_n é $\mathcal{S}_1(1,1,0)$. Portanto, para cada $0 < r < 1$ vem,

$$\mathbb{E} \left| \sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\sqrt{n}\log n} \right|^r \leqslant c_4(r,\sigma,\lambda) < \infty, \quad p = 1, \ldots, n$$

para alguma constante $c_4(r,\sigma,\lambda) > 0$ independente de n o que implica,

$$\mathbb{P}\left\{ \left| \sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\sqrt{n}\log n} \right| \geqslant \delta \right\} \leqslant \frac{\mathbb{E} \left| \sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\sqrt{n}\log n} \right|^r}{\delta^r} \leqslant \frac{c_4(r,\sigma,\lambda)}{\delta^r} < 1$$

escolhendo $\delta > \left[c_4(r,\sigma,\lambda)\right]^{1/r}$ fixo. Então, para todo o $\varepsilon > 0$

$$\mathbb{P}\left\{ \max_{1 \leqslant p \leqslant n} \left| \sum_{j=1}^{p} \frac{w_{ji}(n)\mathrm{e}_j}{\sqrt{n}\,\log n} \right| \geqslant \delta + \varepsilon \right\} \leqslant \frac{\frac{c_4(r,\sigma,\lambda)}{\delta^r}}{1 - \frac{c_4(r,\sigma,\lambda)}{\delta^r}} \cdot \left(\frac{\delta}{\varepsilon} \right)^r$$

(conf. Teorema 1.1.7). Efectuado os cálculos anteriores para cada termo,

$$\sum_{j=p}^{m} \frac{w_{ji}(m)\mathrm{e}_j}{\sqrt{m}\log m}, \quad p = 1, \ldots, m$$

com $m = 1, \ldots, n-1$ concluímos que,

$$\mathbb{P}\left\{ \max_{1 \leqslant p \leqslant n} \left| \sum_{j=1}^{p} \frac{w_{ji}(n)\mathrm{e}_j}{\sqrt{n}\,\log n} \right| \geqslant \delta + \varepsilon \right\} \leqslant \frac{\frac{c_4(r,\sigma,\lambda)}{\delta^r}}{1 - \frac{c_4(r,\sigma,\lambda)}{\delta^r}} \cdot \left(\frac{\delta}{\varepsilon} \right)^r, \quad \forall n \in \mathbb{N}.$$

A tese fica estabelecida usando a identidade (3.1.1) e procedendo como na parte final da demonstração do Teorema 3.1.1. ∎

Teorema 3.2.6 *Se no modelo de regressão linear* (3.0.1),

(a) $e_1, e_2, \ldots$ *são i.i.d.* $\mathcal{S}_1(\sigma, \lambda, \mu)$ *com* $\lambda \neq 0$.

(b) $\rho_n = o\left(\dfrac{1}{n \log^2 n}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em momento de ordem* r *com* $r < \alpha$.

Demonstração. A tese é uma consequência da estimativa (3.2.3) e de

$$\mathbb{E}\left|\sum_{j=1}^{n} \frac{w_{ji}(n) e_j}{\sqrt{n}\log n}\right|^r \leqslant c_4(r, \sigma, \lambda) < \infty, \quad \forall n \in \mathbb{N}$$

para cada $i = 1, \ldots, \kappa$ com $c_4(r, \sigma, \lambda) > 0$ independente de n, uma vez que para todo o $n \in \mathbb{N}$,

$$\sum_{j=1}^{n} \frac{w_{ji}(n) e_j}{\sqrt{n}\log n} \sim \mathcal{S}_1\Big(\bar{\sigma}(n,1), \bar{\lambda}(n,1), \bar{\mu}(n,1)\Big)$$

onde $\bar{\sigma}(n,1), \bar{\lambda}(n,1)$ e $\bar{\mu}(n,1)$ estão definidos por (3.2.6) e (3.2.7) verificam,

$$\bar{\sigma}(n,1) \leqslant \sigma, \quad \left|\bar{\lambda}(n,1)\right| \leqslant |\lambda| \quad \text{e} \quad |\bar{\mu}(n,1)| \leqslant \frac{1}{\log 2} + \frac{\sigma\,|\lambda|}{\pi}. \qquad \blacksquare$$

Observação 3.2.2 Comparando as conclusões dos Teoremas 3.2.3 e 3.2.4 com as dos Teoremas 3.2.5 and 3.2.6 conclui-se que podemos tomar qualquer λ à custa da substituição $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$ por $\rho_n = o\left(\frac{1}{n \log^2 n}\right)$, $n \to \infty$.

3.3 Sob parâmetros σ_i, λ_i e μ_i variáveis

O nosso propósito nesta última secção é alcançar a consistência forte do EMQ generalizando os resultados desenvolvidos anteriormente nas Secções 3.1 e 3.2. Concretamente, assumiremos que os erros são apenas independentes e tais que $e_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, \mu_i)$. Até ao final deste capítulo admitiremos, salvo qualquer indicação em contrário, que $\sigma_1, \sigma_2, \ldots$ é uma sucessão positiva, $\boldsymbol{\sigma}_n = (\sigma_1, \ldots, \sigma_n)$ e $\|\boldsymbol{\sigma}_n\|_\infty = \max\limits_{1 \leqslant j \leqslant n} \sigma_j$. De um modo geral, dada a sucessão real $\mu_1, \mu_2, \ldots$ usaremos a seguinte notação $\boldsymbol{\mu}_n = (\mu_1, \ldots, \mu_n)$.

Teorema 3.3.1 *Se no modelo de regressão linear* (3.0.1),

(a) $e_1, e_2, \ldots$ *são independentes tais que* $e_i \sim \mathcal{S}_2(\sigma_i, \lambda_i, 0)$.

(b) $\rho_n = o\left(\|\boldsymbol{\sigma}_n\|_\infty^{-2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Da Proposição 1.2.1 existirá uma base ortonormada $\{\boldsymbol{w}_1(n), \ldots, \boldsymbol{w}_n(n)\}$ de $\mathbb{R}^n$ tal que,

$$\|\boldsymbol{P}_{\mathbb{X}_n} \mathbf{e}_n\|^2 = \Big(S_{n1}(n)\Big)^2 + \ldots + \Big(S_{n\kappa}(n)\Big)^2$$

onde cada termo de $\{\mathrm{S}_{ni}(n),\, n \geqslant 1\}$ está definido por (3.1.2) para todo o $i = 1, \ldots, \kappa$. As propriedades das distribuições estáveis (conf. Proposição 1.4.1 e Proposição 1.4.3) garantem que,

$$\frac{w_{1i}(n)\mathrm{e}_1}{\|\boldsymbol{\sigma}_n\|_\infty} + \ldots + \frac{w_{ni}(n)\mathrm{e}_n}{\|\boldsymbol{\sigma}_n\|_\infty} \sim \mathcal{S}_2\Big(\bar{\sigma}(n,1), *, 0\Big) \tag{3.3.1}$$

onde,

$$\bar{\sigma}(n,1) = \left[\sum_{j=1}^{n}\left(\frac{\sigma_j\,|w_{ji}(n)|}{\|\boldsymbol{\sigma}_n\|_\infty}\right)^2\right]^{\frac{1}{2}}. \tag{3.3.2}$$

Aplicando a desigualdade de Levy (conf. Teorema 1.1.6) a $\displaystyle\sum_{j=1}^{n}\frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty}$ obtém-se,

$$\begin{aligned}\mathbb{P}\left\{\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p}\frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty}\right| \geqslant \varepsilon\right\} &\leqslant 2\,\mathbb{P}\left\{\left|\sum_{j=1}^{n}\frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty}\right| \geqslant \varepsilon\right\} \leqslant \\ &\leqslant \frac{2\,\mathbb{E}\left(\left|\displaystyle\sum_{j=1}^{n}\frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty}\right|\right)}{\varepsilon} \leqslant \frac{2\sqrt{2}\,\bar{\sigma}(n,1)}{\varepsilon\sqrt{\pi}} \leqslant \frac{2\sqrt{2}}{\varepsilon\sqrt{\pi}}\end{aligned}$$

uma vez que $\sum_{j=1}^{n}\left(\frac{\sigma_j|w_{ji}(n)|}{\|\boldsymbol{\sigma}_n\|_\infty}\right)^2 \leqslant 1$ (conf. alínea (a) da Proposição 1.3.4). Efectuando os cálculos anteriores para cada termo $\sum_{j=1}^{m}\frac{w_{ji}(m)\mathrm{e}_j}{\|\boldsymbol{\sigma}_m\|_\infty}$, $m = 1, \ldots, n-1$ vem,

$$\mathbb{P}\left\{\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p}\frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty}\right| \geqslant \varepsilon\right\} \leqslant \frac{2\sqrt{2}}{\varepsilon\sqrt{\pi}}, \quad \forall n \in \mathbb{N}.$$

Escolhendo a subsucessão $\displaystyle\max_{1\leqslant p\leqslant \xi_n}\left|\sum_{j=1}^{p}\frac{w_{ji}(\xi_n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_{\xi_n}\|_\infty}\right|$ de $\displaystyle\max_{1\leqslant p\leqslant n}\sum_{j=1}^{p}\frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty}$ que nos dá,

$$\lim_{n\to\infty}\left(\max_{1\leqslant j\leqslant \xi_n}\left|\sum_{j=1}^{p}\frac{w_{ji}(\xi_n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_{\xi_n}\|_\infty}\right|\right) = \limsup_{n\to\infty}\left(\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p}\frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty}\right|\right)$$

tem-se,

$$\mathbb{P}\left\{\max_{1\leqslant p\leqslant \xi_n}\left|\sum_{j=1}^{p}\frac{w_{ji}(\xi_n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_{\xi_n}\|_\infty}\right| \geqslant \varepsilon\right\} \leqslant \frac{2\sqrt{2}}{\varepsilon\sqrt{\pi}}, \quad \forall n \in \mathbb{N}.$$

Efectuando a passagem ao limite em n surge,

$$\mathbb{P}\left\{\limsup_{n\to\infty}\left(\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p}\frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty}\right|\right) \geqslant \varepsilon\right\} \leqslant \frac{2\sqrt{2}}{\varepsilon\sqrt{\pi}}$$

e fazendo $\varepsilon \to \infty$ conclui-se que $\displaystyle\limsup_{n\to\infty}\left(\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p}\frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty}\right|\right)$ existe finito quase certamente para todo o $i = 1, \ldots, \kappa$. Consequentemente,

$$\limsup_{n\to\infty}\frac{\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2}{\|\boldsymbol{\sigma}_n\|_\infty^2}$$

existe finito quase certamente verificando-se ainda $\limsup\limits_{n\to\infty} \dfrac{\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2}{\|\boldsymbol{\sigma}_n\|_\infty^2} > 0$ quase certamente. A conclusão da demonstração segue da estimativa (3.0.2) e da hipótese sobre o raio espectral ρ_n. ∎

Também a consistência em média quadrática do EMQ é verdadeira sob idêntica hipótese para o raio espectral.

Teorema 3.3.2 *Se no modelo de regressão linear* (3.0.1)*,*

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes tais que* $\mathrm{e}_i \sim \mathcal{S}_2(\sigma_i, \lambda_i, 0)$.

(b) $\rho_n = o\left(\|\boldsymbol{\sigma}_n\|_\infty^{-2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. A estimativa (3.0.2) e $\displaystyle\sum_{j=1}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\|\boldsymbol{\sigma}_n\|_\infty} \sim \mathcal{S}_2\big(\bar{\sigma}(n,1), *, 0\big) = N\Big(0, 2\big(\bar{\sigma}(n,1)\big)^2\Big)$ garantem que,

$$\mathbb{E}\left[\big\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\big\|^2\right] \leqslant \rho_n\, \mathbb{E}\left[\Big(\mathrm{S}_{n1}(n)\Big)^2 + \ldots + \Big(\mathrm{S}_{n\kappa}(n)\Big)^2\right] = 2\kappa\rho_n\, \|\boldsymbol{\sigma}_n\|_\infty^2 \,\big(\bar{\sigma}(n,1)\big)^2 \leqslant 2\kappa\, \rho_n\, \|\boldsymbol{\sigma}_n\|_\infty^2$$

onde $\bar{\sigma}(n,1)$ estão definidos por (3.3.2). ∎

Corolário 3.3.1 *Se no modelo de regressão linear* (3.0.1)*,*

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes tais que* $\mathrm{e}_i \sim \mathcal{S}_2(\sigma_i, \lambda_i, \mu_i)$.

(b) ρ_n *é uma sucessão evanescente tal que* $\rho_n = o\left(\dfrac{1}{\|\boldsymbol{\sigma}_n\|_\infty^2 + \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. O vector do erros $\mathbf{e}_n$ pode ser decomposto em $\mathbf{e}_n = \boldsymbol{\mu}_n + \mathbf{e}_n^*$ em que cada componente e_i^* do vector aleatório $\mathbf{e}_n^*$ tem distribuição $N(0, 2\sigma_i^2)$. Da linearidade da projecção ortogonal $P_{\mathbb{X}_n}$ sobre o espaço imagem $\mathbb{X}_n$ tem-se,

$$\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2 \leqslant 2\rho_n\, \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2 + 2\rho_n\, \|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n^*\|^2. \tag{3.3.3}$$

Do Teorema 3.3.1 obtém-se $\rho_n\, \|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n^*\|^2 \xrightarrow{\text{q.c.}} 0$ e a conclusão da prova segue da hipótese $\rho_n\, \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2 \longrightarrow 0$ quando $n \to \infty$. ∎

Corolário 3.3.2 *Se no modelo de regressão linear* (3.0.1)*,*

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes tais que* $\mathrm{e}_i \sim \mathcal{S}_2(\sigma_i, \lambda_i, \mu_i)$.

(b) ρ_n *é uma sucessão evanescente tal que* $\rho_n = o\left(\dfrac{1}{\|\boldsymbol{\sigma}_n\|_\infty^2 + \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. A conclusão resulta da combinação da estimativa (3.3.3) com o Teorema 3.3.2. ∎

Teorema 3.3.3 *Se no modelo de regressão linear* (3.0.1),

(a) $e_1, e_2, \ldots$ *são independentes tais que* $e_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, 0)$, $0 < \alpha < 2$, $\alpha \neq 1$.

(b) $\rho_n = o\left[\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{1-\frac{2}{\alpha}}\right]$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Já que vimos que,

$$\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2 = \Big(\mathrm{S}_{n1}(n)\Big)^2 + \ldots + \Big(\mathrm{S}_{n\kappa}(n)\Big)^2$$

onde cada termo de $\{\mathrm{S}_{ni}(n),\, n \geqslant 1\}$ está definido por (3.1.2) para todo o $i = 1, \ldots, \kappa$. As propriedades das distribuições estáveis (conf. Proposição 1.4.1 e Proposição 1.4.3) garantem que,

$$\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{\frac{1}{\alpha}-\frac{1}{2}}} \sim \mathcal{S}_\alpha\Big(\bar{\sigma}(\alpha,n,p), \bar{\lambda}(\alpha,n,p), 0\Big), \quad p = 1, \ldots, n \tag{3.3.4}$$

onde,

$$\bar{\sigma}(\alpha,n,p) = \left[\sum_{j=p}^{n} \frac{\big(\sigma_j\,|w_{ji}(n)|\,\big)^\alpha}{\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{1-\frac{\alpha}{2}}}\right]^{\frac{1}{\alpha}} \tag{3.3.5}$$

e

$$\bar{\lambda}(\alpha,n,p) = \frac{\displaystyle\sum_{j=p}^{n} \mathrm{sign}\big(w_{ji}(n)\big)\lambda_i\,|w_{ji}(n)|^\alpha}{\displaystyle\sum_{j=p}^{n} |w_{ji}(n)|^\alpha}. \tag{3.3.6}$$

Escolhendo $r < \alpha$ vem,

$$\mathbb{P}\left\{\left|\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \delta\right\} \leqslant$$

$$\leqslant \frac{1}{\delta^r}\cdot\mathbb{E}\left|\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{\frac{1}{\alpha}-\frac{1}{2}}}\right|^r \leqslant \frac{\Big(\bar{\sigma}(\alpha,n,p)\cdot c_5\big(r, \bar{\lambda}(\alpha,n,p)\big)\Big)^r}{\delta^r}$$

em que a constante $c_5\big(r, \bar{\lambda}(\alpha,n,p)\big)$ é dada por,

$$\Big(c_5\big(r, \bar{\lambda}(\alpha,n,p)\big)\Big)^r = \frac{2^{r-1}\Gamma\left(1-\frac{r}{\alpha}\right)}{r\int_0^\infty t^{-r-1}\sin^2 t\,dt}\left[1+\big(\bar{\lambda}(\alpha,n,p)\big)^2\tan^2\left(\frac{\alpha\pi}{2}\right)\right]^{\frac{r}{2\alpha}}\cdot$$

$$\cdot\cos\left(\frac{r}{\alpha}\arctan\Big(\bar{\lambda}(\alpha,n,p)\tan\left(\frac{\alpha\pi}{2}\right)\Big)\right).$$

Visto que,

$$\sum_{j=p}^{n} \frac{\big(\sigma_j\,|w_{ji}(n)|\,\big)^{\alpha}}{\Big(\sigma_1^{\frac{2\alpha}{2-\alpha}}+\ldots+\sigma_n^{\frac{2\alpha}{2-\alpha}}\Big)^{1-\frac{\alpha}{2}}} \leqslant 1$$

(conf. alínea (b) da Proposição 1.3.4) obtém-se $\bar{\sigma}(\alpha,n,p)\leqslant 1$ e

$$\mathbb{P}\left\{\left|\sum_{j=p}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\Big(\sigma_1^{\frac{2\alpha}{2-\alpha}}+\ldots+\sigma_n^{\frac{2\alpha}{2-\alpha}}\Big)^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \delta\right\} \leqslant$$

$$\leqslant \frac{\Big(\bar{\sigma}(\alpha,n,p)\cdot c_5\big(r,\bar{\lambda}(\alpha,n,p)\big)\Big)^r}{\delta^r} \leqslant \frac{\big(c_6(\alpha,r)\big)^r}{\delta^r} < 1$$

tomando $\delta > c_6(\alpha,r)$ fixo com,

$$\big(c_6(\alpha,r)\big)^r = \frac{2^{r-1}\Gamma\left(1-\frac{r}{\alpha}\right)}{r\int_0^{+\infty} t^{-r-1}\sin^2 t\,dt}\left[1+\tan^2\left(\frac{\alpha\pi}{2}\right)\right]^{\frac{r}{2\alpha}}.$$

Então, para qualquer $\varepsilon>0$ tem-se,

$$\mathbb{P}\left\{\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p} \frac{w_{ji}(n)\mathrm{e}_j}{\Big(\sigma_1^{\frac{2\alpha}{2-\alpha}}+\ldots+\sigma_n^{\frac{2\alpha}{2-\alpha}}\Big)^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \delta+\varepsilon\right\} \leqslant$$

$$\leqslant \frac{1}{1-\left(\frac{c_6(\alpha,r)}{\delta}\right)^r}\,\mathbb{P}\left\{\left|\sum_{j=1}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\Big(\sigma_1^{\frac{2\alpha}{2-\alpha}}+\ldots+\sigma_n^{\frac{2\alpha}{2-\alpha}}\Big)^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \varepsilon\right\}$$

$$\leqslant \frac{1}{1-\left(\frac{c_6(\alpha,r)}{\delta}\right)^r}\cdot\frac{1}{\varepsilon^r}\cdot\mathbb{E}\left|\sum_{j=1}^{n} \frac{w_{ji}(n)\mathrm{e}_j}{\Big(\sigma_1^{\frac{2\alpha}{2-\alpha}}+\ldots+\sigma_n^{\frac{2\alpha}{2-\alpha}}\Big)^{\frac{1}{\alpha}-\frac{1}{2}}}\right|^r$$

$$\leqslant \frac{\left(\frac{c_6(\alpha,r)}{\delta}\right)^r}{1-\left(\frac{c_6(\alpha,r)}{\delta}\right)^r}\cdot\left(\frac{\delta}{\varepsilon}\right)^r.$$

De forma análoga, podemos efectuar os cálculos precedentes para cada termo,

$$\sum_{j=p}^{m} \frac{w_{ji}(m)\mathrm{e}_j}{\Big(\sigma_1^{\frac{2\alpha}{2-\alpha}}+\ldots+\sigma_m^{\frac{2\alpha}{2-\alpha}}\Big)^{\frac{1}{\alpha}-\frac{1}{2}}},\quad p=1,\ldots,m$$

com $m=1,\ldots,n-1$ e concluir que,

$$\mathbb{P}\left\{\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p} \frac{w_{ji}(n)\mathrm{e}_j}{\Big(\sigma_1^{\frac{2\alpha}{2-\alpha}}+\ldots+\sigma_n^{\frac{2\alpha}{2-\alpha}}\Big)^{\frac{1}{\alpha}-\frac{1}{2}}}\right| \geqslant \delta+\varepsilon\right\} \leqslant \frac{\left(\frac{c_6(\alpha,r)}{\delta}\right)^r}{1-\left(\frac{c_6(\alpha,r)}{\delta}\right)^r}\cdot\left(\frac{\delta}{\varepsilon}\right)^r,\quad \forall\in\mathbb{N}.$$

A tese fica estabelecida raciocinando como na parte final da demonstração do Teorema 3.3.1. ∎

A consistência em momento de ordem r também é verdadeira impondo a mesma condição sobre o raio espectral ρ_n.

Teorema 3.3.4 *Se no modelo de regressão linear* (3.0.1),

(a) $e_1, e_2, \ldots$ *são independentes tais que* $e_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, 0)$, $0 < \alpha < 2$, $\alpha \neq 1$.

(b) $\rho_n = o\left[\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{1-\frac{2}{\alpha}}\right]$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em momento de ordem* r *com* $r < \alpha$.

Demonstração. Para cada $i = 1, \ldots, \kappa$ tem-se,

$$\mathbb{E}\left|\sum_{j=1}^{n} \frac{w_{ji}(n)e_j}{\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{\frac{1}{\alpha}-\frac{1}{2}}}\right|^r \leqslant \Big(c_6(\alpha, r)\Big)^r, \quad \forall n \in \mathbb{N}.$$

uma vez que para todo o $n \in \mathbb{N}$,

$$\sum_{j=1}^{n} \frac{w_{ji}(n)e_j}{\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{\frac{1}{\alpha}-\frac{1}{2}}} \sim \mathcal{S}_\alpha\Big(\bar{\sigma}(\alpha, n, 1), \bar{\lambda}(\alpha, n, 1), 0\Big)$$

onde $\bar{\sigma}(\alpha, n, 1)$ e $\bar{\lambda}(\alpha, n, 1)$ são dados por (3.3.5) and (3.3.6). A tese é agora uma consequência da estimativa (3.2.3) e da hipótese (b). ∎

Corolário 3.3.3 *Se no modelo de regressão linear* (3.0.1),

(a) $e_1, e_2, \ldots$ *são independentes tais que* $e_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, \mu_i)$, $0 < \alpha < 2$, $\alpha \neq 1$.

(b) $\rho_n = o\left(\dfrac{1}{\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{\frac{2}{\alpha}-1} + \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. A tese resulta da combinação entre a estimativa (3.3.3) e o Teorema 3.3.3. ∎

Corolário 3.3.4 *Se no modelo de regressão linear* (3.0.1),

(a) $e_1, e_2, \ldots$ *são independentes tais que* $e_i \sim \mathcal{S}_\alpha(\sigma_i, \lambda_i, \mu_i)$, $0 < \alpha < 2$, $\alpha \neq 1$.

(b) $\rho_n = o\left(\dfrac{1}{\left(\sigma_1^{\frac{2\alpha}{2-\alpha}} + \ldots + \sigma_n^{\frac{2\alpha}{2-\alpha}}\right)^{\frac{2}{\alpha}-1} + \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em momento de ordem* r *com* $r < \alpha$.

Demonstração. A tese é uma consequência da estimativa (3.3.3) e do Teorema 3.3.4. ∎

Os resultados previamente expostos são verdadeiros nalgumas situações particulares do caso $\alpha = 1$. Com efeito, quando $\mathrm{e}_i \sim \mathcal{S}_1(\sigma_i, 0, 0)$ os cálculos realizados nos Teoremas 3.3.3 e 3.3.4 permanecem verdadeiros passo a passo (conf. Corolário 1.4.1). Além disso, a estimativa (3.3.3) continua válida garantindo o,

Teorema 3.3.5 *Se no modelo de regressão linear* (3.0.1)*,*

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes tais que* $\mathrm{e}_i \sim \mathcal{S}_1(\sigma_i, 0, \mu_i)$.

(b) $\rho_n = o\left(\dfrac{1}{\sigma_1^2 + \ldots + \sigma_n^2 + \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Teorema 3.3.6 *Se no modelo de regressão linear* (3.0.1)*,*

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes tais que* $\mathrm{e}_i \sim \mathcal{S}_1(\sigma_i, 0, \mu_i)$.

(b) $\rho_n = o\left(\dfrac{1}{\sigma_1^2 + \ldots + \sigma_n^2 + \|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em momento de ordem* r *com* $r < 1$.

Os próximos resultados cobrem as restantes situações do caso $\alpha = 1$.

Teorema 3.3.7 *Se no modelo de regressão linear* (3.0.1)*,*

(a) $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes tais que* $\mathrm{e}_i \sim \mathcal{S}_1(\sigma_i, \lambda_i, \mu_i)$ *com* $\lambda_i \neq 0$.

(b) $\rho_n = o\left(\dfrac{1}{\|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2 + \|\boldsymbol{\sigma}_n\|_\infty \sqrt{n}\log\left(n\,\|\boldsymbol{\sigma}_n\|_\infty\right)}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Fazendo uso da decomposição $\mathbf{e}_n = \boldsymbol{\mu}_n + \mathbf{e}_n^*$ onde $\boldsymbol{\mu}_n = (\mu_1, \ldots, \mu_n)$ e $\mathbf{e}_n^* = (\mathrm{e}_1^*, \ldots, \mathrm{e}_n^*)$ com $\mathrm{e}_i^* \sim \mathcal{S}_1(\sigma_i, \lambda_i, 0)$ ficamos aptos para utilizar a estimativa (3.3.3). As propriedades das distribuições estáveis (conf. Proposição 1.4.1 e Proposição 1.4.3) asseguram que, para todo o $i = 1, \ldots, \kappa$

$$\sum_{j=p}^{n} w_{ji}(n)\mathrm{e}_j^* \sim \mathcal{S}_1\Big(\sigma^*(n,p), \lambda^*(n,p), \mu^*(n,p)\Big), \quad p = 1, \ldots, n$$

onde,

$$\sigma^*(n,p) = \sum_{j=p}^{n} \sigma_j \left|w_{ji}(n)\right|, \quad \lambda^*(n,p) = \frac{\displaystyle\sum_{j=p}^{n} \operatorname{sign}\big(w_{ji}(n)\big)\lambda_j\sigma_j \left|w_{ji}(n)\right|}{\displaystyle\sum_{j=p}^{n} \sigma_j \left|w_{ji}(n)\right|} \tag{3.3.7}$$

e

$$\mu^*(n,p) = \frac{2}{\pi}\sum_{j=p}^{n} \sigma_j \lambda_j w_{ji}(n) \log\Big(|w_{ji}(n)|^{-1}\Big). \tag{3.3.8}$$

Para qualquer $i = 1, \ldots, \kappa$ tem-se ainda,

$$\sum_{j=p}^{n} \frac{w_{ji}(n) \mathrm{e}_j^*}{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)} \sim \mathcal{S}_1\Big(\bar{\sigma}(n,p), \bar{\lambda}(n,p), \bar{\mu}(n,p)\Big), \quad p = 1, \ldots, n$$

com,

$$\begin{aligned} \bar{\sigma}(n,p) &= \frac{\sigma^*(n,p)}{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \left|\log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)\right|}, \\ \bar{\lambda}(n,p) &= \operatorname{sign}\left(\frac{1}{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)}\right) \lambda^*(n,p) \end{aligned}$$

e

$$\bar{\mu}(n,p) = \frac{\mu^*(n,p)}{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)} + \frac{2}{\pi} \cdot \frac{\sigma^*(n,p)\lambda^*(n,p) \log\left(\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \left|\log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)\right|\right)}{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)}.$$

Da Proposição 1.3.4 é fácil verificar que,

$$\bar{\sigma}(n,p) \leqslant \frac{\sqrt{\sigma_1^2 + \ldots + \sigma_n^2}}{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \left|\log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)\right|} \leqslant \frac{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n}}{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \left|\log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)\right|} \leqslant \frac{1}{\left|\log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)\right|} \leqslant c_7$$

para alguma constante $c_7 > 0$ independente de n e da Proposição 1.3.5 vem ainda,

$$|\bar{\mu}(n,p)| \leqslant \frac{1}{\pi} \cdot \left|\frac{\log n}{\log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)}\right| + \frac{2}{\pi} \cdot \left|\frac{\log\left(\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \left|\log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)\right|\right)}{\log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)}\right| \leqslant c_8$$

para alguma constante $c_8 > 0$ independente de n. Pelo Teorema 1.4.2 obtém-se,

$$\sum_{j=p}^{n} \frac{w_{ji}(n) \mathrm{e}_j^*}{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n} \log\left(n \|\boldsymbol{\sigma}_n\|_\infty\right)} \stackrel{\mathrm{d}}{=} \bar{\sigma}(n,p) \mathrm{Z}_n + \frac{2}{\pi} \bar{\lambda}(n,p) \bar{\sigma}(n,p) \log\big(\bar{\sigma}(n,p)\big) + \bar{\mu}(n,p)$$

com,

$$\begin{aligned} \mathrm{Z}_n \stackrel{\mathrm{d}}{=} & \left(\frac{1 + \bar{\lambda}(n,p)}{2}\right) \mathrm{A}_n - \left(\frac{1 - \bar{\lambda}(n,p)}{2}\right) \mathrm{B}_n + \\ & + \left(\frac{1 + \bar{\lambda}(n,p)}{\pi}\right) \log\left(\frac{1 + \bar{\lambda}(n,p)}{2}\right) - \left(\frac{1 - \bar{\lambda}(n,p)}{\pi}\right) \log\left(\frac{1 - \bar{\lambda}(n,p)}{2}\right) \end{aligned}$$

onde $\{\mathrm{A}_n,\ n \geqslant 1\}$ e $\{\mathrm{B}_n,\ n \geqslant 1\}$ representam duas sucessões de v.a.'s tais que A_n e B_n são independentes com distribuição comum $\mathcal{S}_\alpha(1,1,0)$ para todo o $n \in \mathbb{N}$. Deste modo, para cada $i = 1, \ldots, \kappa$ e qualquer $0 < r < 1$ tem-se,

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{\|\boldsymbol{\sigma}_n\|_\infty \sqrt{n}\log\left(n\|\boldsymbol{\sigma}_n\|_\infty\right)}\sum_{j=p}^{n} w_{ji}(n)\mathrm{e}_j^*\right|^r = \\
&\quad = \mathbb{E}\left|\bar{\sigma}(n,p)\mathrm{Z}_n + \frac{2}{\pi}\bar{\lambda}(n,p)\bar{\sigma}(n,p)\log\left(\bar{\sigma}(n,p)\right) + \bar{\mu}(n,p)\right|^r \\
&\quad \leqslant \mathbb{E}\left(\bar{\sigma}(n,p)\left|\mathrm{Z}_n\right| + \frac{2}{\pi}\left|\bar{\lambda}(n,p)\right|\left|\bar{\sigma}(n,p)\log\left(\bar{\sigma}(n,p)\right)\right| + \left|\bar{\mu}(n,p)\right|\right)^r \\
&\quad \leqslant \mathbb{E}\left(c_7\left|\mathrm{Z}_n\right| + \frac{2C_{\max}}{\pi} + c_8\right)^r \\
&\quad \leqslant c_9\mathbb{E}\left|\mathrm{Z}_n\right|^r + \left(\frac{2C_{\max}}{\pi} + c_8\right)^r, \quad p = 1,\ldots,n.
\end{aligned}$$

em que $c_{\max} = \max\left(e^{-1}, c_7\log c_7\right)$. Analogamente,

$$\begin{aligned}
\mathbb{E}\left(\left|\mathrm{Z}_n\right|^r\right) &= \mathbb{E}\left|\left(\frac{1+\bar{\lambda}(n,p)}{2}\right)\mathrm{A}_n - \left(\frac{1-\bar{\lambda}(n,p)}{2}\right)\mathrm{B}_n + \right. \\
&\left. + \left(\frac{1+\bar{\lambda}(n,p)}{\pi}\right)\log\left(\frac{1+\bar{\lambda}(n,p)}{2}\right) - \left(\frac{1-\bar{\lambda}(n,p)}{\pi}\right)\log\left(\frac{1-\bar{\lambda}(n,p)}{2}\right)\right|^r \\
&\leqslant \mathbb{E}\left[\frac{1+\left|\bar{\lambda}(n,p)\right|}{2}\left|\mathrm{A}_n\right| + \frac{1+\left|\bar{\lambda}(n,p)\right|}{2}\left|\mathrm{B}_n\right| + \right. \\
&\left. + \frac{2}{\pi}\left|\frac{1+\bar{\lambda}(n,p)}{2}\cdot\log\left(\frac{1+\bar{\lambda}(n,p)}{2}\right)\right| + \frac{2}{\pi}\left|\frac{1-\bar{\lambda}(n,p)}{\pi}\log\left(\frac{1-\bar{\lambda}(n,p)}{2}\right)\right|\right]^r \\
&\leqslant \mathbb{E}\left(\left|\mathrm{A}_n\right| + \left|\mathrm{B}_n\right| + \frac{4}{e\pi}\right)^r \\
&\leqslant \mathbb{E}\left|\mathrm{A}_n\right|^r + \mathbb{E}\left|\mathrm{B}_n\right|^r + \left(\frac{4}{e\pi}\right)^r \\
&\leqslant 2c_{10} + \left(\frac{4}{e\pi}\right)^r
\end{aligned}$$

com c_{10} independente de n uma vez que a distribuição de A_n e B_n é $\mathcal{S}_1(1,1,0)$. Então, para cada $0 < r < 1$ vem,

$$\mathbb{E}\left|\sum_{j=p}^{n}\frac{w_{ji}(n)\mathrm{e}_j^*}{\|\boldsymbol{\sigma}_n\|_\infty\sqrt{n}\log\left(n\|\boldsymbol{\sigma}_n\|_\infty\right)}\right|^r \leqslant c_{11}(r) < \infty, \quad p = 1,\ldots,n$$

para alguma constante $c_{11}(r) > 0$ independente de n o que implica,

$$\mathbb{P}\left\{\left|\sum_{j=p}^{n}\frac{w_{ji}(n)\mathrm{e}_j^*}{\|\boldsymbol{\sigma}_n\|_\infty\sqrt{n}\log\left(n\|\boldsymbol{\sigma}_n\|_\infty\right)}\right| \geqslant \delta\right\} \leqslant \frac{c_{11}(r)}{\delta^r} < 1, \qquad p = 1,\ldots,n$$

fixando $\delta > \big[c_{11}(r)\big]^{1/r}$. Para todo o $\varepsilon > 0$,

$$\mathbb{P}\left\{\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p}\frac{w_{ji}(n)\mathrm{e}_j^*}{\|\boldsymbol{\sigma}_n\|_\infty\sqrt{n}\log\left(n\|\boldsymbol{\sigma}_n\|_\infty\right)}\right|\geqslant\delta+\varepsilon\right\}\leqslant$$

$$\leqslant\frac{1}{1-\frac{c_{11}(r)}{\delta^r}}\cdot\frac{\mathbb{E}\left|\sum_{j=p}^{n}\frac{w_{ji}(n)\mathrm{e}_j^*}{\|\boldsymbol{\sigma}_n\|_\infty\sqrt{n}\log\left(n\|\boldsymbol{\sigma}_n\|_\infty\right)}\right|^r}{\varepsilon^r}\leqslant\frac{\frac{c_{11}(r)}{\delta^r}}{1-\frac{c_{11}(r)}{\delta^r}}\cdot\left(\frac{\delta}{\varepsilon}\right)^r$$

(conf. Teorema 1.1.7). Realizando os cálculos anteriores para cada termo,

$$\sum_{j=p}^{m}\frac{w_{ji}(m)\mathrm{e}_j^*}{\|\boldsymbol{\sigma}_m\|_\infty\sqrt{m}\log\left(m\|\boldsymbol{\sigma}_m\|_\infty\right)},\quad p=1,\ldots,m$$

com $m=1,\ldots,n-1$ concluímos que,

$$\mathbb{P}\left\{\max_{1\leqslant p\leqslant n}\left|\sum_{j=1}^{p}\frac{w_{ji}(n)\mathrm{e}_j^*}{\|\boldsymbol{\sigma}_n\|_\infty\sqrt{n}\log\left(n\|\boldsymbol{\sigma}_n\|_\infty\right)}\right|\geqslant\delta+\varepsilon\right\}\leqslant\frac{\frac{c_{11}(r)}{\delta^r}}{1-\frac{c_{11}(r)}{\delta^r}}\cdot\left(\frac{\delta}{\varepsilon}\right)^r,\quad\forall n\in\mathbb{N}.$$

A tese fica estabelecida através da estimativa (3.3.3) e efectuando um raciocínio análogo ao da parte final da demonstração do Teorema 3.3.1. ∎

Teorema 3.3.8 *Se no modelo de regressão linear* (3.0.1),

(a) $\mathrm{e}_1,\mathrm{e}_2,\ldots$ *são independentes tais que* $\mathrm{e}_i\sim\mathcal{S}_1(\sigma_i,\lambda_i,\mu_i)$ *com* $\lambda_i\neq0$.

(b) $\rho_n=o\left(\dfrac{1}{\|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2+\|\boldsymbol{\sigma}_n\|_\infty\sqrt{n}\log\left(n\|\boldsymbol{\sigma}_n\|_\infty\right)}\right)$, $n\to\infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em momento de ordem* r *com* $r<1$.

Demonstração. A tese é uma consequência das estimativas (3.3.3), (3.2.3) e

$$\mathbb{E}\left|\sum_{j=1}^{n}\frac{w_{ji}(n)\mathrm{e}_j^*}{\|\boldsymbol{\sigma}_n\|_\infty\sqrt{n}\log\left(n\|\boldsymbol{\sigma}_n\|_\infty\right)}\right|^r\leqslant c_{11}(r)<\infty\ \text{ para todo o }\ i=1,\ldots,\kappa$$

com $c_{11}(r)>0$ uma constante independente de n e $\mathrm{e}_i^*\sim\mathcal{S}_1(\sigma_i,\lambda_i,0)$. ∎

Observação 3.3.1 Se $\limsup_{n\to\infty}\sigma_n=\sigma\in]0,\infty[$ então o seguinte melhoramento na hipótese (b) dos Teoremas 3.3.7 e 3.3.8 pode ser feito,

$$\rho_n=o\left(\frac{1}{\|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^2+\sqrt{n}\log n}\right),\quad n\to\infty.$$

Observação 3.3.2 Observemos que se $\sigma_n\longrightarrow0$ e $\mu_n\longrightarrow\mu\in\mathbb{R}$ quando $n\to\infty$ então $\mathrm{e}_n\xrightarrow{\mathbb{P}}\mu$ porque $\mathbb{E}\Big[\exp\left(i\mathrm{e}_nt\right)\Big]\longrightarrow\mathbb{E}\Big[\exp\left(i\mu t\right)\Big]$ quando $n\to\infty$. Por outro lado, se ρ_n for uma sucessão evanescente tal que

$$\rho_n=o\Big(\|\boldsymbol{P}_{\mathbb{X}_n}\boldsymbol{\mu}_n\|^{-2}\Big),\quad n\to\infty\quad\text{e}\quad\sum_{n=1}^{\infty}\sigma_n\left|\log\left(\min\{\sigma_n,0.5\}\right)\right|<\infty$$

então a série $\sum_{n=1}^{\infty} \mathrm{e}_n^*$ converge absolutamente quase certamente (conf. Teorema 1.4.6) e a consistência forte do EMQ fica assegurada uma vez que,

$$\left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|^2 \leqslant 2\rho_n \, \left\| \boldsymbol{P}_{\mathbb{X}_n} \boldsymbol{\mu}_n \right\|^2 + 2\rho_n \, \left\| \boldsymbol{P}_{\mathbb{X}_n} \mathbf{e}_n^* \right\|^2 \leqslant 2\rho_n \, \left\| \boldsymbol{P}_{\mathbb{X}_n} \boldsymbol{\mu}_n \right\|^2 + 2\rho_n \, \left\| \mathbf{e}_n^* \right\|^2 .$$

Exemplo 3.3.1

Sejam $\mathrm{e}_1, \mathrm{e}_2, \ldots$ são independentes tais que $\mathrm{e}_i \sim \mathcal{S}_1(\sigma_i, \lambda_i, \mu_i)$ com $\lambda_i \neq 0$. Se $\sigma_n = \delta n^{-\delta} \log n$ $\left(0 < \delta \leqslant \frac{1}{2}\right)$ então $\sigma_n - \sigma_{n+1} \sim \delta^2 n^{-1-\delta} \log n$, $n \to \infty$ o que implica,

$$\sum_{j=n_0}^{n} (\sigma_j - \sigma_{j+1}) \frac{\sqrt{j} \log j}{2} \sim \frac{\delta^2}{1-2\delta} n^{\frac{1}{2}-\delta} \log^2 n, \quad n \to \infty$$

para $0 < \delta < \frac{1}{2}$ e $\sum_{j=n_0}^{n} (\sigma_j - \sigma_{j+1}) \frac{\sqrt{j} \log j}{2} \sim \frac{\log^3 n}{24}$, $n \to \infty$ para $\delta = \frac{1}{2}$. Considerando,

$$\mathbb{H} = \left\{ (t_1, \ldots, t_n) \in \mathbb{R}^n : \; t_1 > 0, \ldots, t_n > 0, \;\; t_1^2 + \ldots + t_n^2 = 1 \right\}$$

obtém-se, usando a transformação de Abel (conf. Secção 1.3),

$$\max_{\mathbb{H}} \sum_{j=1}^{n} \sigma_j t_j \log \left(\frac{1}{t_j} \right) = O\left(n^{\frac{1}{2}-\delta} \log^2 n \right), \quad n \to \infty$$

para $0 < \delta < \frac{1}{2}$ e

$$\max_{\mathbb{H}} \sum_{j=1}^{n} \sigma_j t_j \log \left(\frac{1}{t_j} \right) = O\big(\log^3 n \big), \quad n \to \infty$$

para $\delta = \frac{1}{2}$. Visto que,

$$\sqrt{\sigma_1^2 + \ldots + \sigma_n^2} \sim \frac{\delta}{\sqrt{1-2\delta}} n^{\frac{1}{2}-\delta} \log n, \quad n \to \infty$$

para $0 < \delta < \frac{1}{2}$ e $\sqrt{\sigma_1^2 + \ldots + \sigma_n^2} \sim \dfrac{\sqrt{\log^3 n}}{2\sqrt{3}}$, $n \to \infty$ para $\delta = \frac{1}{2}$ podemos escrever,

$$\rho_n = O\left(\frac{1}{\left\| \boldsymbol{P}_{\mathbb{X}_n} \boldsymbol{\mu}_n \right\|^2 + n^{\frac{1}{2}-\delta} \log^2 n} \right), \quad n \to \infty$$

para $0 < \delta < \frac{1}{2}$ ou

$$\rho_n = O\left(\frac{1}{\left\| \boldsymbol{P}_{\mathbb{X}_n} \boldsymbol{\mu}_n \right\|^2 + \log^3 n} \right), \quad n \to \infty$$

para $\delta = \frac{1}{2}$ porque $\bar{\sigma}(n,p) \leqslant c_{12}$ e $|\bar{\mu}(n,p)| \leqslant c_{13}$ para n suficientemente grande. Naturalmente, se $\sigma_n = \delta n^{-\delta} \log n$ $\left(\frac{1}{2} < \delta \leqslant 1\right)$ então,

$$\max_{\mathbb{H}} \sum_{j=1}^{n} \sigma_j t_j \log \left(\frac{1}{t_j} \right) = O(1), \quad n \to \infty$$

e $\sqrt{\sigma_1^2 + \ldots + \sigma_n^2} = O(1)$, $n \to \infty$. Deste modo, $\sigma^*(n,p) = O(1)$, $|\mu^*(n,p)| = O(1)$, $n \to \infty$ (conf. (3.3.7) e (3.3.8)) e podemos assumir apenas $\rho_n = o\left(\left\| \boldsymbol{P}_{\mathbb{X}_n} \boldsymbol{\mu}_n \right\|^{-2} \right)$, $n \to \infty$.

Capítulo 4

Uma abordagem geométrica

O nosso propósito neste capítulo será analisar a consistência do EMQ supondo que os erros se distribuem radialmente, ou seja, que o vector erros $\mathbf{e}_n$ tem simetria radial (conf. Definição 1.1.14). A grande novidade residirá no facto de não impormos quaisquer condições de independência (na prática associadas geralmente à não transacção de informação) ou idêntica distribuição para a sucessão aleatória $\mathrm{e}_1, \mathrm{e}_2, \ldots$ conseguindo-se, deste modo, abranger um vasto leque de situações em que as observações são feitas. Trabalharemos apenas sob a hipótese de simetria radial do vector dos erros e usaremos coordenadas polares generalizadas para criar outras variáveis aleatórias que nos conduzirão à consistência forte do EMQ. Refira-se, que a radialidade dos erros é condição suficiente para que se tenha a independência das novas variáveis aleatórias resultantes da transformação para coordenadas polares generalizadas do vector dos erros. Por outro lado, desenvolveremos um interessante resultado que consiste na identificação da distribuição de uma variável aleatória criada por factorização e que nos permitirá também obter a consistência forte do EMQ.

Consideremos o modelo de regressão linear,

$$\mathbf{Y}_n = \boldsymbol{X}_n \boldsymbol{\beta} + \mathbf{e}_n \tag{4.0.1}$$

e indiquemos, mais uma vez, o raio espectral da matriz $\left(\boldsymbol{X}_n^T \boldsymbol{X}_n\right)^{-1}$ por ρ_n.

O nosso interesse será estudar o vector aleatório dos erros $\mathbf{e}_n$ sob o ponto de vista geométrico e neste sentido, assumiremos ao longo deste capítulo que $\mathbf{e}_n$ é absolutamente contínuo com simetria radial i.e. que a sua função de densidade conjunta é dada por,

$$f_{\mathrm{e}_1,\ldots,\mathrm{e}_n}(x_1, \ldots, x_n) = g(r), \qquad r = \sqrt{x_1^2 + \ldots + x_n^2}$$

onde g é alguma função não-negativa (conf. Definição 1.1.14). Deste modo, é com toda a naturalidade que surge a transformação de $\mathbf{e}_n$ expresso formalmente em coordenadas cartesianas para coordenadas polares generalizadas. Consideremos pois a aplicação,

$$\mathbb{R}^n \ni (x_1, \ldots, x_n) \mapsto (r, \theta_1, \ldots, \theta_{n-1}) \in]0, \infty[\times]0, \pi[\times \ldots \times]0, \pi[\times]0, 2\pi[$$

definida por,

$$\begin{cases} x_1 = r\cos\theta_1 \\ x_2 = r\sin\theta_1\cos\theta_2 \\ \qquad\vdots \\ x_{n-1} = r\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-2}\cos\theta_{n-1} \\ x_n = r\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-1} \end{cases}$$

(conf. Secção 1.3) cujo o jacobiano é,

$$\mathrm{Jac} = r^{n-1}(\sin\theta_1)^{n-2}(\sin\theta_2)^{n-3}\ldots\sin\theta_{n-2} \geqslant 0.$$

Ao transformamos o vector dos erros $\mathbf{e}_n = (\mathrm{e}_1,\ldots,\mathrm{e}_n)$ usando coordenadas polares generalizadas criamos novas variáveis aleatórias à custa das v.a.'s $\mathrm{e}_1,\ldots,\mathrm{e}_n$, a saber: a **variável aleatória radial** $\mathrm{R}_n = \sqrt{\mathrm{e}_1^2+\ldots+\mathrm{e}_n^2}$ do vector $\mathbf{e}_n$ e os **ângulos aleatórios ao centro** $\Theta_1,\ldots,\Theta_{n-1}$. A função de densidade conjunta de $(\mathrm{R}_n,\Theta_1,\ldots,\Theta_{n-1})$ será então dada por,

$$f_{\mathrm{R}_n,\Theta_1,\ldots,\Theta_{n-1}}(r,\theta_1,\ldots,\theta_{n-1}) = g(r)r^{n-1}(\sin\theta_1)^{n-2}(\sin\theta_2)^{n-3}\ldots\sin\theta_{n-2} \tag{4.0.2}$$

onde g é uma função não-negativa. No próximo resultado provaremos que as novas variáveis aleatórias $\mathrm{R}_n,\Theta_1,\ldots,\Theta_{n-1}$ são (mutuamente) independentes.

Proposição 4.0.1 *Se o vector dos erros* $\mathbf{e}_n$ *é absolutamente contínuo com simetria radial então as variáveis aleatórias* $\mathrm{R}_n,\Theta_1,\ldots,\Theta_{n-2},\Theta_{n-1}$ *são independentes, respectivamente, com densidades*

$$\begin{cases} f_0(r) = \frac{n\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}\,g(r)r^{n-1}, \quad r>0 \\ f_1(\theta) = \frac{(\sin\theta)^{n-2}}{\int_0^\pi(\sin t)^{n-2}\,dt}, \quad 0<\theta<\pi \\ \qquad\vdots \\ f_{n-2}(\theta) = \frac{\sin\theta}{2}, \quad 0<\theta<\pi \\ f_{n-1}(\theta) = \frac{1}{2\pi}, \quad 0<\theta<2\pi. \end{cases}$$

Demonstração. Pondo,

$$f_1(\theta) = \frac{(\sin\theta)^{n-2}}{\int_0^\pi(\sin t)^{n-2}\,dt}, \quad 0<\theta<\pi$$
$$f_2(\theta) = \frac{(\sin\theta)^{n-3}}{\int_0^\pi(\sin t)^{n-3}\,dt}, \quad 0<\theta<\pi$$
$$\vdots$$
$$f_{n-2}(\theta) = \frac{\sin\theta}{2}, \quad 0<\theta<\pi$$
$$f_{n-1}(\theta) = \frac{1}{2\pi}, \quad 0<\theta<2\pi$$

vem,

$$f_1(\theta_1)\ldots f_{n-1}(\theta_{n-1}) = \frac{\Gamma\left(\frac{n}{2}+1\right)}{n\,\pi^{n/2}}(\sin\theta_1)^{n-2}(\sin\theta_2)^{n-3}\ldots\sin\theta_{n-2}.$$

porque a área de superfície da esfera unitária $\mathbb{S}^{n-1}$ de $\mathbb{R}^n$ é

$$\frac{n\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} = \int_{\mathbb{S}^{n-1}} d\boldsymbol{s} =$$
$$= \int_0^{\pi}\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi} (\sin t_1)^{n-2}(\sin t_2)^{n-3}\ldots\sin t_{n-2}\, dt_1\, dt_2\ldots dt_{n-2}\, dt_{n-1}$$

Considerando a função $f_0(r) = \frac{n\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}\, g(r) r^{n-1}$ obtém-se,

$$f_{\mathrm{R}_n,\Theta_1,\ldots,\Theta_{n-1}}(r,\theta_1,\ldots,\theta_{n-1}) = f_0(r) f_1(\theta_1)\ldots f_{n-1}(\theta_{n-1})$$

o que mostra a independência (mútua) de $\mathrm{R}_n, \Theta_1, \ldots, \Theta_{n-1}$ e ainda que as densidades de R_n, Θ_1, $\Theta_2, \ldots, \Theta_{n-2}, \Theta_{n-1}$ são, respectivamente, $f_0, f_1, f_2, \ldots, f_{n-2}, f_{n-1}$ (conf. alínea (b) do Corolário 1.1.1). ∎

Denotemos por,

$$f_{\mathrm{R}_n}(r) = \frac{n\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}\, g(r) r^{n-1}, \quad r > 0 \tag{4.0.3}$$

a densidade da variável aleatória radial R_n que depende da função g não-negativa. Dos ângulos aleatórios ao centro podemos formar o **vector aleatório dos ângulos ao centro**

$$\boldsymbol{\Theta}_{n-1} = (\Theta_1, \ldots, \Theta_{n-1})$$

cuja função de densidade conjunta é (conf. alínea (a) do Corolário 1.1.1),

$$f_{\boldsymbol{\Theta}_{n-1}}(\theta_1,\ldots,\theta_{n-1}) = \frac{\Gamma\left(\frac{n}{2}+1\right)}{n\,\pi^{n/2}}\, (\sin\theta_1)^{n-2}(\sin\theta_2)^{n-3}\ldots\sin\theta_{n-2} \tag{4.0.4}$$

que, como se pode constatar, não depende da função g.

Da factorização,

$$\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2 = \frac{\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2}{\|\mathbf{e}_n\|^2}\ \|\mathbf{e}_n\|^2 = \mathrm{Z}_n\ \|\mathbf{e}_n\|^2 = \mathrm{Z}_n\,\mathrm{R}_n^2 \tag{4.0.5}$$

onde $\mathbb{X}_n = \mathrm{Im}(\boldsymbol{X}_n)$ e

$$\mathrm{Z}_n = \frac{\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2}{\|\mathbf{e}_n\|^2} \tag{4.0.6}$$

somos conduzidos ao importante resultado,

Proposição 4.0.2 *A variável aleatória* $\mathrm{Z}_n = \frac{\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\|^2}{\|\mathbf{e}_n\|^2}$ *é limitada.*

Demonstração. Utilizando a fórmula de Pitágoras tem-se,

$$0 \leqslant \mathrm{Z}_n = \frac{\left\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\right\|^2}{\left\|\mathbf{e}_n\right\|^2} = \frac{\left\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\right\|^2}{\left\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\right\|^2 + \left\|\boldsymbol{P}_{\mathbb{X}_n^{\perp}}\mathbf{e}_n\right\|^2} \leqslant 1$$

o que conclui a prova. ∎

Já sabemos que do facto de $\operatorname{car}(\boldsymbol{X}_n) = \kappa$ e da ausência de multicolinearidade quase-exacta da matriz $\boldsymbol{X}_n$ o EMQ de $\boldsymbol{\beta}$ é expresso por,

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1} \boldsymbol{X}_n^T \mathbf{e}_n \tag{4.0.7}$$

sendo obtido por substituição do vector das observações $\mathbf{Y}_n = \boldsymbol{X}_n \boldsymbol{\beta} + \mathbf{e}_n$ em (1.6.9). Então, apresentamos uma nova versão da estimativa (3.0.2) com a recente variável aleatória Z_n cuja distribuição está perfeitamente definida como veremos na próxima secção.

Proposição 4.0.3 *No modelo de regressão linear* (3.0.1) *tem-se,*

$$\left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|^2 \leqslant \rho_n \, \mathrm{Z}_n \, \mathrm{R}_n^2 \tag{4.0.8}$$

Demonstração. Da estimativa (3.0.2) obtém-se imediatamente

$$\left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|^2 \leqslant \rho_n \, \mathrm{Z}_n \, \mathrm{R}_n^2$$

usando a factorização (4.0.5). ∎

O resultado que se segue é uma extensão da identidade (3.1.1) às novas v.a.'s $\mathrm{R}_n, \Theta_1, \ldots, \Theta_{n-1}$.

Proposição 4.0.4 *Existe uma base ortonormada* $\{\boldsymbol{w}_1(n), \ldots, \boldsymbol{w}_n(n)\}$ *de* $\mathbb{R}^n$ *tal que,*

$$\begin{aligned} \mathrm{Z}_n = \big\langle \boldsymbol{w}_1(n), &(\cos\Theta_1, \sin\Theta_1 \cos\Theta_2, \ldots, \sin\Theta_1 \ldots \sin\Theta_{n-1}) \big\rangle^2 + \\ &+ \ldots + \big\langle \boldsymbol{w}_\kappa(n), (\cos\Theta_1, \sin\Theta_1 \cos\Theta_2, \ldots, \sin\Theta_1 \ldots \sin\Theta_{n-1}) \big\rangle^2, \quad n \geqslant \kappa. \end{aligned}$$

Demonstração. Aplicando a Proposição 1.2.1 a $\boldsymbol{A} = \boldsymbol{X}_n$ e $\boldsymbol{x} = \mathbf{e}_n$ existirá uma base ortonormada $\{\boldsymbol{w}_1(n), \ldots, \boldsymbol{w}_n(n)\}$ de $\mathbb{R}^n$ tal que a v.a. $\|\boldsymbol{P}_{\mathbb{X}_n} \mathbf{e}_n\|^2$ é expressa por,

$$\|\boldsymbol{P}_{\mathbb{X}_n} \mathbf{e}_n\|^2 = \langle \boldsymbol{w}_1(n), \mathbf{e}_n \rangle^2 + \ldots + \langle \boldsymbol{w}_\kappa(n), \mathbf{e}_n \rangle^2$$

e tese fica estabelecida efectuando a transformação para coordenadas polares generalizadas de $\mathbf{e}_n = (\mathrm{e}_1, \ldots, \mathrm{e}_n)$

$$\begin{cases} \mathrm{e}_1 = \mathrm{R}_n \cos\Theta_1 \\ \mathrm{e}_2 = \mathrm{R}_n \sin\Theta_1 \cos\Theta_2 \\ \qquad\vdots \\ \mathrm{e}_{n-1} = \mathrm{R}_n \sin\Theta_1 \sin\Theta_2 \ldots \sin\Theta_{n-2} \cos\Theta_{n-1} \\ \mathrm{e}_n = \mathrm{R}_n \sin\Theta_1 \sin\Theta_2 \ldots \sin\Theta_{n-1} \end{cases}$$

tendo em conta que $\|\mathbf{e}_n\|^2 = \mathrm{R}_n^2$. ∎

Observação 4.0.3 Observemos que da Proposição 4.0.1 conclui-se também a independência de R_n e Z_n uma vez que, pela Proposição 4.0.4, Z_n apenas depende dos ângulos ao centro $\Theta_1, \ldots, \Theta_{n-1}$.

4.1 A distribuição de Z_n

Sob a hipótese da simetria radial do vector dos erros $\mathbf{e}_n$, vimos na secção anterior que podemos utilizar coordenadas polares generalizadas para transformar $(\mathrm{e}_1, \ldots, \mathrm{e}_n)$ em $(\mathrm{R}_n, \Theta_{n-1}, \ldots, \Theta_{n-1})$ cuja função de densidade conjunta é dada por (4.0.2). Da Proposição 4.0.4 é ainda fácil verificar que a variável aleatória Z_n depende unicamente dos ângulos ao centro $\Theta_1, \ldots, \Theta_{n-1}$ e portanto a sua função densidade de probabilidade não dependerá da função g. Então, para acharmos a densidade da variável aleatória Z_n podemos supor qualquer função g na densidade conjunta do vector dos erros. Em particular, escolhendo

$$g(r) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{r^2}{2}}$$

somos conduzidos à seguinte densidade conjunta para $\mathbf{e}_n$,

$$f_{\mathrm{e}_1,\ldots,\mathrm{e}_n}(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{x_1^2+\ldots+x_n^2}{2}}$$

o que corresponde à situação em que o vector aleatório dos erros tem densidade normal multivariada, i.e. $\mathbf{e}_n \sim N(\mathbf{0}, \boldsymbol{I})$. Consequentemente, as componentes $\mathrm{e}_1, \ldots, \mathrm{e}_n$ são independentes tendo cada uma distribuição normal standard unidimensional (conf. o Teorema 1.2.8 e o Teorema 1.2.9). Pelo teorema de Cochran (conf. Teorema 1.2.10) as variáveis aleatórias $\left\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\right\|^2$ e $\left\|\boldsymbol{P}_{\mathbb{X}_n^\perp}\mathbf{e}_n\right\|^2$ onde $\mathbb{X}_n = \mathrm{Im}(\boldsymbol{X}_n)$ são independentes tendo-se,

$$\left\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\right\|^2 \sim \chi^2(\kappa) \quad \text{e} \quad \left\|\boldsymbol{P}_{\mathbb{X}_n^\perp}\mathbf{e}_n\right\|^2 \sim \chi^2(n-\kappa).$$

Proposição 4.1.1 *Sejam* $\mathrm{X}_1, \ldots, \mathrm{X}_m$ *variáveis aleatórias independentes com densidades* $f_{\mathrm{X}_i}(x_i)$ $(i = 1, \ldots, m)$ *e* $\mathrm{Y}_1, \ldots, \mathrm{Y}_m$ *as variáveis aleatórias definidas por* $\mathrm{Y}_j = \mathrm{X}_1 + \ldots + \mathrm{X}_j$, $j = 1, \ldots, m$. *Então a função de densidade conjunta do vector aleatório* $(\mathrm{Y}_1, \ldots, \mathrm{Y}_m)$ *é,*

$$f_{\mathrm{Y}_1,\ldots,\mathrm{Y}_m}(y_1, \ldots, y_m) = f_{\mathrm{X}_1}(y_1)\, f_{\mathrm{X}_2}(y_2 - y_1) \ldots f_{\mathrm{X}_m}(y_m - y_{m-1}).$$

Demonstração. O sistema de m equações lineares e m incógnitas $x_1, \ldots, x_m$

$$\begin{cases} x_1 = y_1 \\ x_1 + x_2 = y_2 \\ \qquad \vdots \\ x_1 + x_2 + \ldots + x_m = y_m \end{cases}$$

tem uma única solução dada por,

$$\begin{cases} x_1 = y_1 \\ x_2 = y_2 - y_1 \\ \qquad \vdots \\ x_m = y_m - y_{m-1} \end{cases}.$$

Então, a densidade conjunta do vector aleatório $(\mathrm{Y}_1, \ldots, \mathrm{Y}_m)$ será,

$$f_{\mathrm{Y}_1,\ldots,\mathrm{Y}_m}(y_1, \ldots, y_m) = \frac{f_{\mathrm{X}_1,\ldots,\mathrm{X}_m}(x_1, \ldots, x_m)}{|\mathrm{Jac}(x_1, \ldots, x_m)|} = f_{\mathrm{X}_1}(y_1) \cdot f_{\mathrm{X}_2}(y_2 - y_1) \ldots f_{\mathrm{X}_m}(y_m - y_{m-1})$$

pois as variáveis aleatórias $\mathrm{X}_1, \ldots, \mathrm{X}_m$ são independentes e o jacobiano da transformação,

$$\begin{cases} h_1(x_1, \ldots, x_m) = x_1 \\ h_2(x_1, \ldots, x_m) = x_1 + x_2 \\ \qquad \vdots \\ h_m(x_1, \ldots, x_m) = x_1 + x_2 + \ldots + x_m \end{cases}$$

é,

$$\mathrm{Jac}(x_1, \ldots, x_m) = \det\left(\begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \ldots & \frac{\partial h_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \ldots & \frac{\partial h_m}{\partial x_m} \end{bmatrix}\right) = \det\left(\begin{bmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 1 & \ldots & 1 \end{bmatrix}\right) = 1. \qquad \blacksquare$$

Definindo as variáveis aleatórias $\mathrm{X}_1 = \left\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\right\|^2$, $\mathrm{X}_2 = \left\|\boldsymbol{P}_{\mathbb{X}_n^\perp}\mathbf{e}_n\right\|^2$, $\mathrm{Y}_1 = \left\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\right\|^2$ e $\mathrm{Y}_2 = \left\|\mathbf{e}_n\right\|^2$ concluímos da Proposição 4.1.1 que a função de densidade conjunta do par aleatório $(\mathrm{Y}_1, \mathrm{Y}_2)$ é $f_{\mathrm{Y}_1,\mathrm{Y}_2}(x, y) = f_{\mathrm{X}_1}(x)\, f_{\mathrm{X}_2}(y - x)$. Deste modo, a densidade da variável aleatória Z_n é,

$$\begin{aligned} f_{\mathrm{Z}_n}(z) &= \int_{-\infty}^{\infty} |\, t\,|\; f_{\mathrm{Y}_1,\mathrm{Y}_2}(zt, t)\, dt \\ &= \int_{-\infty}^{\infty} |\, t\,|\; f_{\mathrm{X}_1}(zt)\, f_{\mathrm{X}_2}(t - zt)\, dt \\ &= \frac{1}{\Gamma\left(\frac{\kappa}{2}\right)\Gamma\left(\frac{n-\kappa}{2}\right)} z^{\frac{\kappa}{2}-1}(1-z)^{\frac{n-\kappa}{2}-1} \int_0^{\infty} \frac{1}{2^{\frac{n}{2}}}\, t^{\frac{n}{2}-1} e^{-\frac{t}{2}}\, dt \\ &= \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{\kappa}{2}\right)\Gamma\left(\frac{n-\kappa}{2}\right)} z^{\frac{\kappa}{2}-1}(1-z)^{\frac{n-\kappa}{2}-1} \end{aligned}$$

sempre que $0 < z < 1$ e $f_{\mathrm{Z}_n}(z) = 0$ se caso contrário. Portanto, a variável aleatória Z_n tem distribuição beta de parâmetros $\left(\frac{\kappa}{2}, \frac{n-\kappa}{2}\right)$.

Convém destacar que podíamos ter seguido um caminho alternativo ao exposto acima e evitado a Proposição 4.1.1. Com efeito, a variável aleatória W_n definida por,

$$\mathrm{W}_n = \frac{\dfrac{\left\|\boldsymbol{P}_{\mathbb{X}_n}\mathbf{e}_n\right\|^2}{\kappa}}{\dfrac{\left\|\boldsymbol{P}_{\mathbb{X}_n^\perp}\mathbf{e}_n\right\|^2}{n-\kappa}}$$

tem distribuição de Snedecor com graus de liberdade κ e $(n - \kappa)$ i.e.

$$f_{\mathrm{W}_n}(w) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{\kappa}{2}\right)\Gamma\left(\frac{n-\kappa}{2}\right)} \left(\frac{\kappa}{n-\kappa}\right)^{\frac{\kappa}{2}} w^{\frac{\kappa}{2}-1} \left(1 + \frac{\kappa w}{n-\kappa}\right)^{-\frac{n}{2}}, \quad w > 0.$$

Deste modo, a densidade da variável aleatória $\mathrm{V}_n = \dfrac{\kappa}{n-\kappa}\,\mathrm{W}_n$ será,

$$f_{\mathrm{V}_n}(v) = \frac{\Gamma\left(\dfrac{n}{2}\right)}{\Gamma\left(\dfrac{\kappa}{2}\right)\Gamma\left(\dfrac{n-\kappa}{2}\right)}\, v^{\frac{\kappa}{2}-1}\,(1+v)^{-\frac{n}{2}}, \quad v > 0.$$

Fazendo uso da transformação $u = \dfrac{v}{v+1}$ obtemos a identidade integral,

$$\int_t^\infty v^{a-1}\,(1+v)^{-a-b}\,dv = \int_{\frac{t}{t+1}}^1 u^{a-1}\,(1-u)^{b-1}\,du, \quad t > 0$$

e tomando $a = \frac{\kappa}{2}$ e $b = \frac{n-\kappa}{2}$ vem,

$$F_{\mathrm{V}_n}(t) = F_{\mathrm{U}_n}\left(\frac{t}{t+1}\right), \quad t > 0 \tag{4.1.1}$$

onde U_n é uma variável aleatória com distribuição beta de parâmetros $\left(\frac{\kappa}{2}, \frac{n-\kappa}{2}\right)$. Adicionalmente, visto que

$$\mathrm{Z}_n \leqslant \mathrm{V}_n \tag{4.1.2}$$

temos ainda que,

$$\mathbb{P}(\mathrm{Z}_n > t) \leqslant \mathbb{P}(\mathrm{V}_n > t) = \mathbb{P}\left(\mathrm{U}_n > \frac{t}{t+1}\right), \qquad t > 0 \tag{4.1.3}$$

e podemos utilizar a distribuição de U_n para obter majorações de $\mathbb{P}(\mathrm{Z}_n > t)$.

4.2 A consistência do estimador

O objectivo central desta secção é obter a consistência forte do EMQ. Mas antes, estabeleçamos alguns resultados auxiliares.

Lema 4.2.1 *Para qualquer* $0 \leqslant \alpha < 1$ *existe* $n_0 \in \mathbb{N}$ *tal que,*

$$\mathbb{P}\left(\mathrm{Z}_n > \frac{\varepsilon}{n^\alpha}\right) < \frac{\Gamma\left(\dfrac{n}{2}\right)}{\Gamma\left(\dfrac{\kappa}{2}\right)\Gamma\left(\dfrac{n-\kappa}{2}\right)}\left(\frac{\varepsilon}{n^\alpha}\right)^{\frac{\kappa}{2}-1}\left(1-\frac{\varepsilon}{n^\alpha}\right)^{\frac{n-\kappa}{2}}, \qquad \varepsilon > 0$$

para todo o $n \geqslant n_0$.

Demonstração. Se $\kappa \geqslant 2$ e $n > \kappa + 2$ então Z_n tem em

$$z_0 = \frac{\kappa-2}{n-4}$$

a única moda (se $\kappa = 1$ e $n > 3$ então $f_{\mathrm{Z}_n}(z)$ é monótona decrescente em $]0,1[$). Visto que para todo o $0 < \alpha < 1$,

$$\lim_{n\to\infty} \frac{\dfrac{\kappa-2}{n-4}}{\dfrac{\varepsilon}{n^\alpha}} = 0, \qquad \varepsilon > 0$$

existirá uma ordem $n_0 \in \mathbb{N}$ tal que,

$$z_0 = \frac{\kappa - 2}{n - 4} < \frac{\varepsilon}{n^\alpha}, \qquad \forall n \geqslant n_0, \quad \varepsilon > 0.$$

Deste modo, como a densidade $f_{Z_n}(z)$ é monótona decrescente em $]z_0, 1[$ podemos escrever,

$$\mathbb{P}\left(Z_n > \frac{\varepsilon}{n^\alpha}\right) < \left(1 - \frac{\varepsilon}{n^\alpha}\right) f_{Z_n}\left(\frac{\varepsilon}{n^\alpha}\right), \quad \forall n \geqslant n_0, \quad \varepsilon > 0$$

o que estabelece a tese. ∎

Lema 4.2.2 *Para qualquer* $0 < \alpha < 1$ *tem-se,*

$$\lim_{n\to\infty} n^\xi \left(1 - \frac{\varepsilon}{n^\alpha}\right)^{\frac{n-\kappa}{2}} = 0, \qquad \xi \in \mathbb{R}, \quad \varepsilon > 0.$$

Demonstração. Dado $\xi \in \mathbb{R}$ tem-se para todo o $\varepsilon > 0$,

$$\begin{aligned} n^\xi \left(1 - \frac{\varepsilon}{n^\alpha}\right)^{\frac{n-\kappa}{2}} &= e^{\log\left(n^\xi \left(1-\frac{\varepsilon}{n^\alpha}\right)^{\frac{n}{2}}\right)} \left(1 - \frac{\varepsilon}{n^\alpha}\right)^{-\frac{\kappa}{2}} \\ &= e^{\xi \log n + \frac{n^{1-\alpha}}{2} \log\left(1-\frac{\varepsilon}{n^\alpha}\right)^{n^\alpha}} \left(1 - \frac{\varepsilon}{n^\alpha}\right)^{-\frac{\kappa}{2}} \\ &= e^{n^{1-\alpha}\left(\frac{\xi \log n}{n^{1-\alpha}} + \frac{1}{2} \log\left(1-\frac{\varepsilon}{n^\alpha}\right)^{n^\alpha}\right)} \left(1 - \frac{\varepsilon}{n^\alpha}\right)^{-\frac{\kappa}{2}} \end{aligned}$$

e como $0 < \alpha < 1$ vem,

$$\lim_{n\to\infty} \log\left(1 - \frac{\varepsilon}{n^\alpha}\right)^{n^\alpha} = -\varepsilon \qquad \text{e} \qquad \lim_{n\to\infty} \frac{\xi \log n}{n^{1-\alpha}} = 0$$

donde,

$$\lim_{n\to\infty} n^\xi \left(1 - \frac{\varepsilon}{n^\alpha}\right)^{\frac{n-\kappa}{2}} = 0.$$ ∎

Vamos, de seguida, enunciar dois importantes resultados com influência directa e decisiva para consistência forte do estimador dos mínimos quadrados.

Proposição 4.2.1 *Para todo o* $0 < \alpha < 1$ *tem-se,*

$$\mathbb{P}\left(\limsup_{n\to\infty} \left\{Z_n > \frac{1}{m\, n^\alpha}\right\}\right) = 0, \qquad m = 1, 2, \ldots$$

Demonstração. Visto que,

$$\lim_{n\to\infty} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-\kappa}{2}\right) n^{\frac{\kappa}{2}}} = 2^{-\frac{\kappa}{2}}$$

(ver [60]) tem-se do Lema 4.2.2 com $\xi = \alpha\left(1-\dfrac{\kappa}{2}\right)+s$ $(s>1)$ e $\varepsilon = \dfrac{1}{m}$ $(m=1,2,\ldots)$

$$\lim_{n\to\infty}\frac{\dfrac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{\kappa}{2}\right)\Gamma\left(\frac{n-\kappa}{2}\right)}\left(\dfrac{1}{m\,n^{\alpha}}\right)^{\frac{\kappa}{2}-1}\left(1-\dfrac{1}{m\,n^{\alpha}}\right)^{\frac{n-\kappa}{2}}}{\dfrac{1}{n^{s}}}=$$

$$=\frac{1}{m^{\frac{\kappa}{2}-1}\,\Gamma\left(\frac{\kappa}{2}\right)\,2^{\frac{\kappa}{2}}}\lim_{n\to\infty} n^{\alpha+\frac{\kappa}{2}(1-\alpha)+s}\left(1-\frac{1}{m\,n^{\alpha}}\right)^{\frac{n-\kappa}{2}}=0.$$

Então o Lema 4.2.1 garante que,

$$\lim_{n\to\infty}\frac{\mathbb{P}\left(\mathrm{Z}_n>\dfrac{1}{m\,n^{\alpha}}\right)}{\dfrac{1}{n^{s}}}\leqslant\lim_{n\to\infty}\frac{\dfrac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{\kappa}{2}\right)\Gamma\left(\frac{n-\kappa}{2}\right)}\left(\dfrac{1}{m\,n^{\alpha}}\right)^{\frac{\kappa}{2}-1}\left(1-\dfrac{1}{m\,n^{\alpha}}\right)^{\frac{n-\kappa}{2}}}{\dfrac{1}{n^{s}}}=0$$

e consequentemente,

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\mathrm{Z}_n>\frac{1}{m\,n^{\alpha}}\right)<\infty,\qquad m=1,2,\ldots$$

porque a série $\sum_{n=1}^{\infty}\frac{1}{n^{s}}$ $(s>1)$ é convergente. A conclusão surge do primeiro lema de Borel-Cantelli (conf. Teorema 1.1.1). ∎

Observação 4.2.1 Optando por seguir o caminho alternativo descrito no final da secção anterior ter-se-ia, de um modo natural, que dado um qualquer $0<\alpha<1$ existiria $n_0\in\mathbb{N}$ tal que,

$$\mathbb{P}\left(\mathrm{U}_n>\frac{\varepsilon}{n^{\alpha}+\varepsilon}\right)<\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{\kappa}{2}\right)\Gamma\left(\frac{n-\kappa}{2}\right)}\left(\frac{\varepsilon}{n^{\alpha}}\right)^{\frac{\kappa}{2}-1}\left(1-\frac{\varepsilon}{n^{\alpha}+\varepsilon}\right)^{\frac{n-\kappa}{2}},\quad \varepsilon>0$$

para todo o $n\geqslant n_0$. Além disso,

$$\lim_{n\to\infty} n^{\xi}\left(1-\frac{\varepsilon}{n^{\alpha}+\varepsilon}\right)^{\frac{n-\kappa}{2}}=0,\qquad \xi\in\mathbb{R},\quad \varepsilon>0$$

o que permitiria obter a conclusão da Proposição 4.2.1 à custa desigualdade (4.1.2).

Proposição 4.2.2 *Para qualquer* $0<\alpha<1$ *tem-se* $n^{\alpha}\,\mathrm{Z}_n\xrightarrow{\text{q.c.}}0$.

Demonstração. De acordo com a Proposição 4.2.1 temos,

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{|n^{\alpha}\,\mathrm{Z}_n|>\frac{1}{m}\right\}\right)=\mathbb{P}\left(\limsup_{n\to\infty}\left\{\mathrm{Z}_n>\frac{1}{m\,n^{\alpha}}\right\}\right)=0$$

para cada $m=1,2,\ldots$ o que implica (ver [18]),

$$\mathbb{P}\left(\lim_{n\to\infty}n^{\alpha}\,\mathrm{Z}_n=0\right)=1$$

como pretendíamos. ∎

Apresentemos agora os primeiros resultados sobre a consistência forte do EMQ.

Teorema 4.2.1 *Se o vector dos erros* $\mathbf{e}_n$ *no modelo de regressão linear* (4.0.1) *é absolutamente contínuo com simetria radial e para algum* $0 < \alpha < 1$,

$$\exists \mathrm{C} = \mathrm{C}(\omega) > 0\colon \quad \limsup_{n\to\infty} \left(n^{-\alpha}\rho_n \, \mathrm{R}_n^2\right) \leqslant \mathrm{C}(\omega) \quad q.c. \tag{4.2.1}$$

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. A tese é uma consequência da estimativa (4.0.8) e da Proposição 4.2.2. ∎

Corolário 4.2.1 *Se o vector dos erros* $\mathbf{e}_n$ *no modelo de regressão linear* (4.0.1) *é absolutamente contínuo com simetria radial, a variável aleatória radial* $\mathrm{R}_n \xrightarrow{\text{q.c.}} \mathrm{R}_\infty$ *e* $\rho_n = O(n^\alpha)$, $n \to \infty$ *para algum* $0 < \alpha < 1$ *então* $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Das hipóteses assumidas resulta que o termo $n^{-\alpha}\rho_n \, \mathrm{R}_n^2$ é limitado quase certamente donde a conclusão vem do Teorema 4.2.1. ∎

A consistência em média quadrática do EMQ pode ser obtida à custa da independência de R_n e Z_n (conf. Observação 4.0.3) impondo, adicionalmente, condições sobre $\mathbb{E}\left(\mathrm{R}_n^2\right)$.

Teorema 4.2.2 *Se o vector dos erros* $\mathbf{e}_n$ *no modelo de regressão linear* (4.0.1) *é absolutamente contínuo com simetria radial,* $\mathbb{E}\left(\mathrm{R}_n^2\right) = O\left(n^p\right)$, $n \to \infty$ *e* $\rho_n = o\left(n^{1-p}\right)$, $n \to \infty$ *então* $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. De acordo com a estimativa (4.0.8) tem-se,

$$\mathbb{E}\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2\right) \leqslant \rho_n \, \mathbb{E}\left(\mathrm{R}_n^2\right) \, \mathbb{E}\left(\mathrm{Z}_n\right) \leqslant c_1 \, \rho_n \, n^p \, \frac{\kappa}{n} = \kappa \, c_1 \cdot \frac{\rho_n}{n^{1-p}} = o(1)$$

quando $n \to \infty$ e a tese fica estabelecida. ∎

Mais geralmente, a independência das variáveis aleatórias R_n e Z_n conduzem-nos à consistência em momento de ordem s do EMQ.

Teorema 4.2.3 *Se o vector dos erros* $\mathbf{e}_n$ *no modelo de regressão linear* (4.0.1) *é absolutamente contínuo com simetria radial,* $\mathbb{E}\left(\mathrm{R}_n^s\right) = O\left(n^p\right)$, $n \to \infty$ *e* $\rho_n = o\left(n^{1-\frac{2p}{s}}\right)$, $n \to \infty$ *então* $\widetilde{\boldsymbol{\beta}}$ *é consistente em momento de ordem* s.

Demonstração. Dado $s > 0$ a estimativa (4.0.8) produz,

$$\mathbb{E}\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^s\right) \leqslant {\rho_n}^{\frac{s}{2}} \cdot \mathbb{E}\left(\mathrm{R}_n^s\right) \, \mathbb{E}\left({\mathrm{Z}_n}^{\frac{s}{2}}\right) = {\rho_n}^{\frac{s}{2}} \cdot \mathbb{E}\left(\mathrm{R}_n^s\right) \frac{\Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{s+\kappa}{2}\right)}{\Gamma\left(\frac{\kappa}{2}\right) \Gamma\left(\frac{n+s}{2}\right)} = o(1)$$

quando $n \to \infty$ porque $\Gamma\left(\dfrac{n+s}{2}\right) \sim \left(\dfrac{n}{2}\right)^{s/2} \Gamma\left(\dfrac{n}{2}\right)$, $n \to \infty$. ∎

Observação 4.2.2 A consistência do EMQ em momento de ordem s permanece válida se $\mathbf{e}_n$ for absolutamente contínuo com simetria radial e $\rho_n = o\left(\dfrac{n}{[\mathbb{E}(\mathrm{R}_n^s)]^{2/s}}\right)$, $n \to \infty$.

Convém agora destacar uma situação particular da simetria radial relativa ao caso em que os erros são independentes (não-degenerados). Neste cenário, como veremos já de seguida, somos automaticamente conduzidos à normalidade de $\mathrm{e}_1, \ldots, \mathrm{e}_n$.

Teorema 4.2.4 *Se as variáveis aleatórias* $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes e o vector dos erros* $\mathbf{e}_n$ *tem simetria radial,*

$$f_{\mathrm{e}_1,\ldots,\mathrm{e}_n}(x_1,\ldots,x_n) = g(r), \qquad r = \sqrt{x_1^2+\ldots+x_n^2}$$

com função g *diferenciável então cada erro* $\mathrm{e}_1, \mathrm{e}_2, \ldots, \mathrm{e}_n$ *tem distribuição normal com valor médio nulo e igual variância.*

Demonstração. De $f_{\mathrm{e}_1,\ldots,\mathrm{e}_n}(x_1,\ldots,x_n) = g(r)$, $r = \sqrt{x_1^2+\ldots+x_n^2}$ e

$$f_{\mathrm{e}_1,\ldots,\mathrm{e}_n}(x_1,\ldots,x_n) = f_{\mathrm{e}_1}(x_1)\ldots f_{\mathrm{e}_n}(x_n)$$

tem-se,

$$g\left(\sqrt{x_1^2+\ldots+x_n^2}\right) = f_{\mathrm{e}_1}(x_1)\ldots f_{\mathrm{e}_n}(x_n). \tag{4.2.2}$$

Visto que,

$$\frac{\partial g(r)}{\partial x_1} = \frac{dg(r)}{dr}\frac{\partial r}{\partial x_1} \quad \text{e} \quad \frac{\partial r}{\partial x_1} = \frac{x_1}{r}$$

concluímos, derivando (4.2.2) em ordem a x_1, que

$$\frac{x_1}{r}\,g'(r) = f'_{\mathrm{e}_1}(x_1) f_{\mathrm{e}_2}(x_2)\ldots f_{\mathrm{e}_n}(x_n).$$

Dividindo ambos os membros por $x_1\,g(r) = x_1\,f_{\mathrm{e}_1}(x_1) f_{\mathrm{e}_2}(x_2)\ldots f_{\mathrm{e}_n}(x_n)$ obtém-se,

$$\frac{1}{r}\frac{g'(r)}{g(r)} = \frac{1}{x_1}\frac{f'_{\mathrm{e}_1}(x_1)}{f_{\mathrm{e}_1}(x_1)} \tag{4.2.3}$$

O lado direito de (4.2.3) é independente de $x_2,\ldots,x_n$ e o lado esquerdo de (4.2.3) é uma função de $r = \sqrt{x_1^2+\ldots+x_n^2}$. Então ambos os lados são necessariamente independentes de $x_1, x_2, \ldots, x_n$ donde,

$$\frac{1}{r}\frac{g'(r)}{g(r)} = c_2 = \text{constante}.$$

Consequentemente,

$$\frac{d\log g(r)}{dr} = c_2 r \quad \Longrightarrow \quad g(r) = c_3\,e^{\frac{c_2\,r^2}{2}}$$

e a condição de simetria radial produz

$$f_{\mathrm{e}_1,\ldots,\mathrm{e}_n}(x_1,\ldots,x_n) = g\left(\sqrt{x_1^2+\ldots+x_n^2}\right) = c_3\,e^{c_2\cdot\frac{x_1^2+\ldots+x_n^2}{2}}.$$

Do facto de $f_{\mathrm{e}_1,\ldots,\mathrm{e}_n}$ ser função de densidade conjunta obtém-se $c_3 = \left(-\frac{c_2}{2\pi}\right)^{\frac{n}{2}}$ vindo,

$$f_{\mathrm{e}_1,\ldots,\mathrm{e}_n}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\left(-\frac{1}{c_2}\right)^{\frac{n}{2}}} e^{-\frac{1}{2}\boldsymbol{x}\,\mathrm{diag}(-c_2,\ldots,-c_2)\,\boldsymbol{x}^T}, \quad \boldsymbol{x} \in \mathbb{R}^n$$

e portanto, as variáveis aleatórias $\mathrm{e}_1, \mathrm{e}_2, \ldots \mathrm{e}_n$ são normais com média zero e variância $\sigma^2 = -\dfrac{1}{c_2}$ (conf. Teorema 1.2.9). ∎

Teorema 4.2.5 *Se as variáveis aleatórias* $\mathrm{e}_1, \mathrm{e}_2, \ldots$ *são independentes com densidades* $f_{\mathrm{e}_1}, f_{\mathrm{e}_2}, \ldots$ *contínuas e o vector dos erros* $\mathbf{e}_n$ *tem simetria radial então cada erro* $\mathrm{e}_1, \mathrm{e}_2, \ldots, \mathrm{e}_n$ *tem distribuição normal com valor médio nulo e igual variância.*

Demonstração. De acordo com as hipóteses tem-se,

$$f_{\mathrm{e}_1}(x_1)\ldots f_{\mathrm{e}_n}(x_n) = g(r), \qquad r^2 = x_1^2 + \ldots + x_n^2. \tag{4.2.4}$$

De (4.2.4) surge,

$$\begin{aligned} f_{\mathrm{e}_1}(x_1) f_{\mathrm{e}_2}(0)\ldots f_{\mathrm{e}_n}(0) &= g\left(|x_1|\right) \\ f_{\mathrm{e}_1}(0) f_{\mathrm{e}_2}(x_2)\ldots f_{\mathrm{e}_n}(0) &= g\left(|x_2|\right) \\ &\vdots \\ f_{\mathrm{e}_1}(0)\ldots f_{\mathrm{e}_{n-1}}(0) f_{\mathrm{e}_n}(x_n) &= g\left(|x_n|\right) \end{aligned}$$

o que implica,

$$g(r) = f_{\mathrm{e}_1}(r) f_{\mathrm{e}_2}(0)\ldots f_{\mathrm{e}_n}(0),\ f_{\mathrm{e}_2}(x_2) = \frac{f_{\mathrm{e}_1}(x_2)}{f_{\mathrm{e}_1}(0)}\, f_{\mathrm{e}_2}(0),\ \ldots\ , f_{\mathrm{e}_n}(x_n) = \frac{f_{\mathrm{e}_1}(x_n)}{f_{\mathrm{e}_1}(0)}\, f_{\mathrm{e}_n}(0).$$

Consequentemente, a equação funcional (4.2.4) pode ser escrita como, $h(t) = \log\left(\dfrac{f_{\mathrm{e}_1}(t)}{f_{\mathrm{e}_1}(0)}\right)$,

$$h(x_1) + \ldots + h(x_n) = h(r), \qquad r^2 = x_1^2 + \ldots + x_n^2. \tag{4.2.5}$$

Além disso, fazendo $x_2 = \ldots = x_n = 0$, $x_1 = -x_1$ em (4.2.5), podemos afirmar ainda que $h(x_1) = h(-x_1) = h(|x_1|)$. Portanto, se

$$x_1^2 = y_1^2 + y_2^2,$$

$$h(r) = h(y_1) + h(y_2) + h(x_2) + \ldots + h(x_n), \qquad r^2 = y_1^2 + y_2^2 + x_2^2 + \ldots + x_n^2,$$

tem-se, em geral,

$$h(r) = h(y_1) + \ldots + h(y_m), \qquad r^2 = y_1^2 + \ldots + y_m^2.$$

Escolhendo $m = p^2$ e pondo $y = y_1 = \ldots = y_m$, obtém-se,

$$h(p\,y) = p^2\,h(y) \quad \text{ou} \quad h(p) = p^2\,h(1) \quad \text{para} \quad y = 1.$$

Para $y = \dfrac{q}{p}$ onde q é um inteiro vem,

$$p^2\, h\left(\frac{q}{p}\right) = h\left(p \cdot \frac{q}{p}\right) = h(q) = q^2\, h(1) \quad \text{ou} \quad h\left(\frac{q}{p}\right) = c\left(\frac{q}{p}\right)^2,$$

onde $c_4 = \varphi(1)$. Assim, $\varphi(y) = c_4\, y^2$ para todo o racional y e da hipótese da continuidade de $f_{\mathrm{e}_1}, f_{\mathrm{e}_2}, \ldots$ concluí-se que a relação é válida para todo y. Então,

$$f_{\mathrm{e}_1}(y) = f_{\mathrm{e}_1}(0)\, e^{c_4\, y^2}. \tag{4.2.6}$$

e para esta função ser uma densidade de probabilidade, c_4 deve ser não-negativo e ser escrito como $-\dfrac{1}{2\sigma^2}$. Integrando (4.2.6) em $\mathbb{R}$ e igualando o resultado à unidade tem-se $f_{\mathrm{e}_1}(0) = \dfrac{1}{\sigma\sqrt{2\pi}}$ donde,

$$f_{\mathrm{e}_1}(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{y^2}{2\sigma^2}}, \qquad y \in \mathbb{R}$$

que corresponde à distribuição normal $N(0, \sigma^2)$ com $\mathbb{E}(\mathrm{e}_1) = 0$ e $\mathbb{V}(\mathrm{e}_1) = \sigma^2$. As densidades de $\mathrm{e}_2, \ldots, \mathrm{e}_n$ são também normais $N(0, \sigma^2)$ pois integrando em $\mathbb{R}$ as funções seguintes,

$$f_{\mathrm{e}_2}(x_2) = \frac{f_{\mathrm{e}_1}(x_2)}{f_{\mathrm{e}_1}(0)}\, f_{\mathrm{e}_2}(0),\ \ldots\ , f_{\mathrm{e}_n}(x_n) = \frac{f_{\mathrm{e}_1}(x_n)}{f_{\mathrm{e}_1}(0)}\, f_{\mathrm{e}_n}(0)$$

e igualando os respectivos resultados à unidade produz $f_{\mathrm{e}_2}(0) = \ldots = f_{\mathrm{e}_n}(0) = \dfrac{1}{\sigma\sqrt{2\pi}}$. ■

4.3 A extensão ao caso $\alpha = 1$

Na Secção 4.1 vimos que, se o vector dos erros for absolutamente contínuo com simetria radial então a variável aleatória Z_n tem distribuição beta com parâmetros $\left(\frac{\kappa}{2}, \frac{n-\kappa}{2}\right)$. Então,

$$f_{n\mathrm{Z}_n}(z) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{\kappa}{2}\right)\Gamma\left(\frac{n-\kappa}{2}\right) n^{\frac{\kappa}{2}}}\, z^{\frac{\kappa}{2}-1}\left(1 - \frac{z}{n}\right)^{\frac{n-\kappa}{2}-1}, \qquad 0 < z < n$$

e $f_{n\mathrm{Z}_n}(z) = 0$ se caso contrário. Da Proposição 4.0.4 sabemos que a variável aleatória $n\mathrm{Z}_n$ pode ser expressa por,

$$n\,\mathrm{Z}_n = \Big(\mathrm{S}_{(n-1)1}(n)\Big)^2 + \ldots + \Big(\mathrm{S}_{(n-1)\kappa}(n)\Big)^2$$

onde para cada $i = 1, \ldots, \kappa$,

$$\begin{aligned}\mathrm{S}_{(n-1)i}(n) = \sqrt{n}\cos\Theta_1 w_{1i}(n) + \sqrt{n}\sin\Theta_1\cos\Theta_2 w_{2i}(n) + \ldots + \sqrt{n}\sin\Theta_1 \ldots \\ \sin\Theta_{n-2}\cos\Theta_{n-1} w_{(n-1)i}(n) + \sqrt{n}\sin\Theta_1 \ldots \sin\Theta_{n-2}\sin\Theta_{n-1} w_{ni}(n).\end{aligned}$$

Consideremos pois o arranjo triangular $\big\{\mathrm{T}_{mi}(n),\ n \geqslant 2,\ 1 \leqslant m \leqslant n-1\big\}$ de variáveis aleatórias definido por,

$$\left\{\ \mathrm{T}_{1i}(2) = \sqrt{2}\,\big[\cos\Theta_1 w_{1i}(2) + \sin\Theta_1 w_{2i}(2)\big]\right.$$

$$\begin{cases} \mathrm{T}_{1i}(3) = \sqrt{3}\,\cos\Theta_1 w_{1i}(3) \\ \mathrm{T}_{2i}(3) = \sqrt{3}\,\sin\Theta_1\big[\cos\Theta_2 w_{2i}(3) + \sin\Theta_2 w_{3i}(3)\big] \end{cases}$$

$$\vdots$$

$$\begin{cases} \mathrm{T}_{1i}(n) = \sqrt{n}\,\cos\Theta_1 w_{1i}(n) \\ \mathrm{T}_{2i}(n) = \sqrt{n}\,\sin\Theta_1\cos\Theta_2 w_{2i}(n) \\ \qquad\vdots \\ \mathrm{T}_{(n-2)i}(n) = \sqrt{n}\,\sin\Theta_1\ldots\sin\Theta_{n-3}\cos\Theta_{n-2}w_{(n-2)i}(n) \\ \mathrm{T}_{(n-1)i}(n) = \sqrt{n}\,\sin\Theta_1\ldots\sin\Theta_{n-2}\big[\cos\Theta_{n-1}w_{(n-1)i}(n) + \sin\Theta_{n-1}w_{ni}(n)\big] \end{cases}$$

para todo o $i = 1, \ldots, \kappa$. Então,

$$\begin{aligned} \mathrm{S}_{1i}(2) &= \mathrm{T}_{1i}(2) \\ \mathrm{S}_{2i}(3) &= \mathrm{T}_{1i}(3) + \mathrm{T}_{2i}(3) \\ &\ \ \vdots \\ \mathrm{S}_{(n-1)i}(n) &= \mathrm{T}_{1i}(n) + \mathrm{T}_{2i}(n) + \ldots + \mathrm{T}_{(n-1)i}(n) \end{aligned}$$

para todo o $i = 1, \ldots, \kappa$. Visto que,

$$\begin{aligned} &\mathbb{E}\big[\mathrm{T}_{1i}(2)\big] = 0 \quad \text{q.c.} \\ &\mathbb{E}\big[\mathrm{T}_{2i}(3) \mid \mathrm{T}_{1i}(3)\big] = \sqrt{3}\sin\Theta_1\big[w_{2i}(3)\,\mathbb{E}(\cos\Theta_2) + w_{3i}(3)\,\mathbb{E}(\sin\Theta_2)\big] = 0 \quad \text{q.c.} \\ &\qquad\vdots \\ &\mathbb{E}\big[\mathrm{T}_{(n-1)i}(n) \mid \mathrm{T}_{1i}(n), \ldots, \mathrm{T}_{(n-2)i}(n)\big] = \\ &\qquad = \sqrt{n}\sin\Theta_1\ldots\sin\Theta_{n-2}\big[w_{(n-1)i}(n)\,\mathbb{E}(\cos\Theta_{n-1}) + w_{ni}(n)\,\mathbb{E}(\sin\Theta_{n-1})\big] = 0 \quad \text{q.c.} \end{aligned}$$

podemos aplicar a igualdade extendida de Bienaymé e a desigualdade extendida de Kolmogorov (conf. Teorema 1.1.13 e Teorema 1.1.14) a cada termo da sucessão aleatória $\mathrm{S}_{(n-1)i}(n)$, $n \geqslant 2$ o que produz,[1]

$$\begin{aligned} \mathbb{P}\left\{\max_{1\leqslant j\leqslant 1}|\mathrm{S}_{ji}(2)| \geqslant \varepsilon\right\} &\leqslant \frac{1}{\varepsilon^2}\cdot\mathbb{E}\left[\Big(\mathrm{S}_{1i}(2)\Big)^2\right] \leqslant \frac{\mathbb{E}\big(2\,\mathrm{Z}_2\big)}{\varepsilon^2} = \frac{\kappa}{\varepsilon^2} \\ \mathbb{P}\left\{\max_{1\leqslant j\leqslant 2}|\mathrm{S}_{ji}(3)| \geqslant \varepsilon\right\} &\leqslant \frac{1}{\varepsilon^2}\cdot\mathbb{E}\left[\Big(\mathrm{S}_{2i}(3)\Big)^2\right] \leqslant \frac{\mathbb{E}\big(3\,\mathrm{Z}_3\big)}{\varepsilon^2} = \frac{\kappa}{\varepsilon^2} \\ &\ \ \vdots \\ \mathbb{P}\left\{\max_{1\leqslant j\leqslant n-1}|\mathrm{S}_{ji}(n)| \geqslant \varepsilon\right\} &\leqslant \frac{1}{\varepsilon^2}\cdot\mathbb{E}\left[\Big(\mathrm{S}_{(n-1)i}(n)\Big)^2\right] \leqslant \frac{\mathbb{E}\big(n\,\mathrm{Z}_n\big)}{\varepsilon^2} = \frac{\kappa}{\varepsilon^2} \end{aligned}$$

uma vez que $\mathbb{E}\big(n\mathrm{Z}_n\big) = \kappa$, $\forall n \in \mathbb{N}$. Consequentemente,

$$\mathbb{P}\left\{\max_{1\leqslant j\leqslant n-1}|\mathrm{S}_{ji}(n)| \geqslant \varepsilon\right\} \leqslant \frac{\kappa}{\varepsilon^2}, \quad \forall n \in \mathbb{N}$$

[1] Observemos que $\mathbb{E}\big[\mathrm{S}_{(n-1)i}(n)\big] = 0$, $\forall n \geqslant 2$.

e e escolhendo a subsucessão $\max\limits_{1\leqslant j\leqslant \eta_n-1} |\mathrm{S}_{ji}(\eta_n)|$ de $\max\limits_{1\leqslant j\leqslant n-1} |\mathrm{S}_{ji}(n)|$ que nos dá,

$$\lim_{n\to\infty}\left(\max_{1\leqslant j\leqslant \eta_n-1} |\mathrm{S}_{ji}(\eta_n)|\right) = \limsup_{n\to\infty}\left(\max_{1\leqslant j\leqslant n-1} |\mathrm{S}_{ji}(n)|\right)$$

vem,

$$\mathbb{P}\left\{\max_{1\leqslant j\leqslant \eta_n-1} |\mathrm{S}_{ji}(\eta_n)| \geqslant \varepsilon\right\} \leqslant \frac{\kappa}{\varepsilon^2}, \quad \forall n\in\mathbb{N}.$$

Efectuando a passagem ao limite em n surge,

$$\mathbb{P}\left\{\limsup_{n\to\infty}\left(\max_{1\leqslant j\leqslant n-1} |\mathrm{S}_{ji}(n)|\right) \geqslant \varepsilon\right\} \leqslant \frac{\kappa}{\varepsilon^2}$$

e fazendo $\varepsilon\to\infty$ concluí-se que $\limsup\limits_{n\to\infty}\left(\max\limits_{1\leqslant j\leqslant n-1} |\mathrm{S}_{ji}(n)|\right)$ existe finito quase certamente para todo o $i=1,\ldots,\kappa$. Portanto,

$$\limsup_{n\to\infty}\ n\mathrm{Z}_n$$

existe finito quase certamente. Por outro lado, é fácil verificar que a sucessão de funções $\{f_{n\mathrm{Z}_n},\ n\geqslant 1\}$ converge pontualmente em $\mathbb{R}$ quando $n\to\infty$ i.e.

$$f_{n\mathrm{Z}_n}(z)\longrightarrow f(z) = \begin{cases} \dfrac{1}{2^{\frac{\kappa}{2}}\,\Gamma\left(\dfrac{\kappa}{2}\right)}\, z^{\frac{\kappa}{2}-1}e^{-\frac{z}{2}} & \text{se} \quad z>0 \\ \\ 0 & \text{se} \quad z\leqslant 0 \end{cases}$$

quando $n\to\infty$. Então $n\mathrm{Z}_n \xrightarrow{\mathrm{d}} \chi^2(\kappa)$, ou seja, $n\mathrm{Z}_n$ converge em lei para uma variável aleatória que possui distribuição χ^2 com κ graus de liberdade o que implica $\limsup\limits_{n\to\infty}\ n\mathrm{Z}_n>0$ quase certamente.

Desta forma, podemos realizar melhoramentos nos resultados da anterior Secção 4.2.

Teorema 4.3.1 *Seja $p\geqslant 0$. Se o vector dos erros $\mathbf{e}_n$ no modelo de regressão linear* (4.0.1) *é absolutamente contínuo com simetria radial, $\rho_n = o\left(n^{-p}\right)$, $n\to\infty$ e*

$$\exists\mathrm{C}=\mathrm{C}(\omega)>0\colon \quad \limsup_{n\to\infty}\frac{\mathrm{R}_n^2}{n^{p+1}}\leqslant \mathrm{C}(\omega) \quad \textit{q.c.} \tag{4.3.1}$$

então $\widetilde{\boldsymbol{\beta}}$ é fortemente consistente.

Demonstração. A condição (4.3.1) conduz-nos a,

$$\rho_n\,\mathrm{R}_n^2\,\mathrm{Z}_n = \rho_n\,n^p\cdot\frac{\mathrm{R}_n^2}{n^{p+1}}\cdot n\mathrm{Z}_n \leqslant \mathrm{C}_1(\omega)\,\rho_n\,n^p \quad \text{q.c.}$$

e a conclusão da prova acontece tendo em conta a hipótese sobre o raio espectral ρ_n. ■

Corolário 4.3.1 *Seja $p\geqslant 0$. Se o vector dos erros $\mathbf{e}_n$ no modelo de regressão linear* (4.0.1) *é absolutamente contínuo com simetria radial, $\dfrac{\mathrm{R}_n}{n^p}\xrightarrow{\text{q.c.}}\mathrm{R}_\infty$ e $\rho_n = o\left(n^{1-2p}\right)$, $n\to\infty$ então $\widetilde{\boldsymbol{\beta}}$ é fortemente consistente.*

Demonstração. A hipótese $\dfrac{\mathrm{R}_n}{n^p} \xrightarrow{\text{q.c.}} \mathrm{R}_\infty$ implica $\dfrac{\mathrm{R}_n^2}{n^{2p}} \xrightarrow{\text{q.c.}} \mathrm{R}_\infty^2$ (ver [18] ou [75]) e a tese segue do Teorema 4.3.1. ∎

Completamos esta abordagem geométrica à consistência do EMQ exibindo alguns exemplos de aplicação.

Exemplo 4.3.1 (Distribuição Normal Multivariada)
Supondo,

$$g(r) = \frac{1}{(2\pi\,\sigma^2)^{n/2}}\, e^{-\frac{r^2}{2\,\sigma^2}}, \qquad \sigma > 0$$

então cada componente do vector $\mathbf{e}_n$ tem distribuição $N(0, \sigma^2)$. Então, as componentes $\mathrm{e}_1, \ldots, \mathrm{e}_n$ são i.i.d. para todo o $n \geqslant 2$ e de acordo com a lei forte de Kolmogorov (conf. Observação 1.1.7) vem,

$$\frac{\mathrm{R}_n^2}{n} = \frac{\mathrm{e}_1^2 + \ldots + \mathrm{e}_n^2}{n} \xrightarrow{\text{q.c.}} \sigma^2$$

uma vez que $\mathbb{E}\left(\mathrm{e}_i^2\right) = \sigma^2$, $\forall i$. O Corolário 4.3.1 assegura assim que o EMQ é fortemente consistente ($p = \frac{1}{2}$). Além disso, R_n^2 tem distribuição $\sigma^2\chi^2(n)$ o que implica $\mathbb{E}\left(\mathrm{R}_n^2\right) = \sigma^2 n$ e a consistência em média quadrática do EMQ é verdadeira quando $\rho_n = o(1)$, $n \to \infty$. É de destacar que recuperámos precisamente a mesma conclusão do Teorema 3.1.1.

Exemplo 4.3.2 (Distribuição de Kotz)
Sejam $a, b > 0$ e $n + 2q > 2$. Se,

$$g(r) = \frac{a\,\Gamma\left(\dfrac{n}{2}\right)}{b^{-\frac{2q+n-2}{2a}}\, \pi^{\frac{n}{2}}\, \Gamma\left(\dfrac{2q+n-2}{2a}\right)}\, r^{2q-2} e^{-b\,r^{2a}}$$

então o vector dos erros $\mathbf{e}_n$ tem distribuição de Kotz[2] (ver [84]). Por exemplo, tomando $q = a - \frac{n}{2} + 1$ obtém-se,

$$f_{\mathrm{R}_n^2}(r) = \frac{1}{2\sqrt{r}}\, f_{\mathrm{R}_n}\left(\sqrt{r}\right) = ab\, r^{a-1} e^{-b\,r^a}, \quad r > 0.$$

ou seja, R_n^2 tem distribuição de Weibull-Gnedenko com parâmetros (a, b). Visto que,

$$\mathbb{E}\big(\mathrm{R}_n^2\big) = b^{-\frac{1}{a}}\, \Gamma\left(1 + \frac{1}{a}\right)$$

então a consistência em média quadrática é verdadeira se $\rho_n = o(n)$, $n \to \infty$. A consistência forte do EMQ também é válida se $\rho_n = o(n)$, $n \to \infty$ porque,

$$\sup_{m>n} \mathbb{E}\left|\mathrm{R}_m^2 - \mathrm{R}_n^2\right| = \sup_{m>n} \Big(\mathbb{E}\big(\mathrm{R}_m^2\big) - \mathbb{E}\big(\mathrm{R}_n^2\big)\Big) = 0$$

[2]Note-se que se $q = a = 1$ e $b = \frac{1}{2\sigma^2}$ ($\sigma > 0$) então recuperamos o anterior exemplo da distribuição normal multivariada.

implica $\mathrm{R}_n^2 \xrightarrow{\mathscr{L}_1} \mathrm{R}_\infty$ para alguma v.a. $\mathrm{R}_\infty \in \mathscr{L}_1$. Portanto, $\mathrm{R}_n^2 \xrightarrow{\text{q.c.}} \mathrm{R}_\infty$ uma vez que R_n^2 é uma sucessão monótona não-decrescente e $\mathrm{R}_n^2 \xrightarrow{\mathbb{P}} \mathrm{R}_\infty$ (ver [57]).

Exemplo 4.3.3 (Distribuição Uniforme Multivariada)
Seja $a > 0$ e consideremos,

$$g(r) = \frac{\Gamma\left(\frac{n}{2}+1\right)}{(a\sqrt{\pi})^n}, \qquad r < a$$

que corresponde à situação em que o vector dos erros $\mathbf{e}_n$ está distribuído uniformemente numa bola aberta centrada na origem com raio a. É fácil verificar que R_n tem densidade,

$$f_{\mathrm{R}_n}(r) = \frac{n}{a^n}\, r^{n-1}, \qquad 0 < r < a$$

e distribuição,

$$F_{\mathrm{R}_n}(r) = \begin{cases} 0 & \text{se} \quad r \leqslant 0 \\ \left(\dfrac{r}{a}\right)^n & \text{se} \quad 0 < r < a \\ 1 & \text{se} \quad r \geqslant a \end{cases}.$$

Então,

$$F_{\mathrm{R}_n}(r) \longrightarrow F(r) = \begin{cases} 1 & \text{se} \quad r \geqslant a \\ 0 & \text{se} \quad r < a \end{cases}$$

quando $n \to \infty$ e $\mathrm{R}_n \xrightarrow{\mathbb{P}} a$ donde $\mathrm{R}_n^2 \xrightarrow{\text{q.c.}} a^2$ uma vez que R_n^2 é uma sucessão não-decrescente (ver [57]). Do Corolário 4.3.1 a consistência forte do EMQ fica garantida se $\rho_n = o(n)$, $n \to \infty$. Por outro lado, a variável aleatória R_n^2 tem densidade,

$$f_{\mathrm{R}_n^2}(r) = \frac{n}{2\,a^n}\, r^{\frac{n}{2}-1}, \qquad 0 < r < a^2$$

o que implica $\mathbb{E}\left(\mathrm{R}_n^2\right) = \dfrac{n\,a^2}{n+2}$ e a consistência em média quadrática também fica assegurada se $\rho_n = o(n)$, $n \to \infty$.

Exemplo 4.3.4 (Distribuição t Multivariada)
Seja $q \in \mathbb{N}$ e suponhamos,

$$g(r) = \frac{\Gamma\left(\dfrac{n+q}{2}\right) n^{\frac{n}{2}}}{\Gamma\left(\dfrac{q}{2}\right)(q\pi)^{\frac{n}{2}}} \left(1 + \frac{n}{q}\, r^2\right)^{-\frac{n+q}{2}} \tag{4.3.2}$$

que corresponde à situação em que o vector dos erros $\mathbf{e}_n$ tem distribuição t multivariada com q graus de liberdade, matriz de precisão $\boldsymbol{T} = \mathrm{diag}\big(n, \ldots, n\big)$ e distorção nula. A densidade de R_n^2 será,

$$f_{\mathrm{R}_n^2}(r) = \frac{1}{2\sqrt{r}}\, f_{\mathrm{R}_n}\left(\sqrt{r}\right) = \frac{n\,\pi^{n/2}}{2\,\Gamma\left(\dfrac{n}{2}+1\right)}\, g\left(\sqrt{r}\right)\, r^{\frac{n}{2}-1}, \qquad r > 0$$

donde substituindo a função g dada em (4.3.2) vem,

$$f_{\mathrm{R}_n^2}(r) = \frac{\Gamma\left(\dfrac{n+q}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\,\Gamma\left(\dfrac{q}{2}\right)} \left(\frac{n}{q}\right)^{\frac{n}{2}} r^{\frac{n}{2}-1} \left(1 + \frac{n\,r}{q}\right)^{-\frac{n+q}{2}}, \qquad r > 0.$$

Portanto, R_n^2 tem distribuição de Snedecor com graus de liberdade n e q o que implica,

$$\mathbb{E}\big(\mathrm{R}_n^2\big) = \frac{q}{q-2} \qquad (q > 2).$$

Então, a consistência em média quadrática é verdadeira se $\rho_n = o(n)$, $n \to \infty$. A consistência forte também é válida se $\rho_n = o(n)$, $n \to \infty$: com efeito,

$$\sup_{m>n} \mathbb{E}\left|\mathrm{R}_m^2 - \mathrm{R}_n^2\right| = \sup_{m>n}\Big(\mathbb{E}\big(\mathrm{R}_m^2\big) - \mathbb{E}\big(\mathrm{R}_n^2\big)\Big) = 0$$

implica $\mathrm{R}_n^2 \xrightarrow{\mathscr{L}_1} \mathrm{R}_\infty$ para alguma v.a. $\mathrm{R}_\infty \in \mathscr{L}_1$. Deste modo, $\mathrm{R}_n^2 \xrightarrow{\text{q.c.}} \mathrm{R}_\infty$ uma vez que R_n^2 é uma sucessão não-decrescente e $\mathrm{R}_n^2 \xrightarrow{\mathbb{P}} \mathrm{R}_\infty$ (ver [57]).

Capítulo 5

Em domínios de atracção maximais

Neste capítulo, investigaremos a consistência forte e em média quadrática do EMQ através do comportamento da cauda de probabilidade da distribuição do quadrado dos erros. Efectivamente, assumindo que as variáveis aleatórias $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \dots$ são independentes e identicamente distribuídas, o nosso grande propósito será estabelecer a consistência do EMQ em cada um dos domínios de atracção maximal (conf. Secção 1.5). Acontece que, num cenário de convergência em lei da sucessão centrada e normalizada

$$\frac{\max\left(\mathrm{e}_1^2, \dots, \mathrm{e}_n^2\right) - d_n}{c_n} \qquad (c_n > 0, \;\; d_n \in \mathbb{R})$$

onde a respectiva sucessão de funções de distribuição converge para uma distribuição não-degenerada H, a variável aleatória e^2 pertencerá ao domínio de atracção maximal de H. Além disso, H é do mesmo tipo de uma das seguintes distribuições de valores extremos: Fréchet, Weibull ou Gumbel (conf. Teorema 1.5.2). Refira-se que, em particular, quando as variáveis aleatórias $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \dots$ são não-degeneradas, independentes e identicamente distribuídas tal convergência em lei é verificada quando a variável aleatória e^2 é max-estável, i.e. quando para todo $n \geqslant 2$ é verdadeira a identidade em lei,

$$\max\left(\mathrm{e}_1^2, \dots, \mathrm{e}_n^2\right) \stackrel{\mathrm{d}}{=} c_n\, \mathrm{e}^2 + d_n$$

para algumas constantes $c_n > 0$ e $d_n \in \mathbb{R}$ (conf. Definição 1.5.1). Deste modo, vamos estabelecer a consistência forte e em média quadrática do EMQ quando a distribuição do quadrado dos erros pertence a uma das três grandes classes:

A. Domínio de atracção maximal da distribuição de Fréchet.

B. Domínio de atracção maximal da distribuição de Weibull.

C. Domínio de atracção maximal da distribuição de Gumbel.

Suponhamos então o modelo de regressão linear,

$$\mathbf{Y}_n = \boldsymbol{X}_n \boldsymbol{\beta} + \mathbf{e}_n. \tag{5.0.1}$$

Visto estarmos a admitir que a matriz de modelo $\boldsymbol{X}_n$ tem característica κ e ainda a excluir qualquer hipótese de multicolinearidade quase-exacta da matriz $\boldsymbol{X}_n$ então o EMQ é dado por,

$$\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\boldsymbol{X}_n^T \boldsymbol{X}_n\right)^{-1} \boldsymbol{X}_n^T \mathbf{e}_n$$

e utilizando a estimativa (3.0.2) obtém-se sucessivamente,

$$\begin{aligned}\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2 \leqslant \rho\Big((\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1}\Big) \left\|\boldsymbol{P}_{\Omega_n} \mathbf{e}_n\right\|^2 \underset{\substack{\text{fórmula de}\\ \text{Pitágoras}}}{\leqslant} \rho\Big((\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1}\Big) \left\|\mathbf{e}_n\right\|^2 \leqslant \\ \underset{\left\|\mathbf{e}_n\right\|^2 \leqslant n \max\left(\mathrm{e}_1^2, \ldots, \mathrm{e}_n^2\right)}{\leqslant} \rho\Big((\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1}\Big) n \mathrm{M}_n.\end{aligned} \tag{5.0.2}$$

onde $\mathrm{M}_n = \max\left(\mathrm{e}_1^2, \ldots, \mathrm{e}_n^2\right)$. À semelhança de anteriores capítulos, denotaremos por ρ_n o raio espectral da matriz $\left(\boldsymbol{X}_n^T \boldsymbol{X}_n\right)^{-1}$.

5.1 O domínio de atracção maximal da distribuição Fréchet

Antes expormos os resultados centrais desta secção analisamos, como forma de motivação para o que se vai seguir, um caso concreto de uma cauda de probabilidade.

Teorema 5.1.1 *Se $r > 0$ e no modelo de regressão linear* (5.0.1)*,*

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.*

(b) $\overline{F}_{\mathrm{e}^2}(x) = O\big(x^{-1/r}\big)$, $x \to \infty$.

(c) $\exists s > 2r + 1\colon\ \rho_n = O\big(1/n^s\big)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ é fortemente consistente.

Demonstração. Da estimativa (5.0.2) apenas teremos de mostrar que,

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{\infty} \mathbb{P}\big\{n\, \rho_n\, \mathrm{M}_n > \varepsilon\big\} < \infty$$

(conf. Teorema 1.1.21). Mas isto é imediato porque $\sum_{n=1}^{\infty} n^{\frac{r-s+1}{r}} < \infty$ e

$$\begin{aligned}\mathbb{P}\big\{n\rho_n \mathrm{M}_n > \varepsilon\big\} = 1 - \left(1 - \overline{F}_{\mathrm{e}^2}\left(\frac{\varepsilon}{n\,\rho_n}\right)\right)^n \leqslant 1 - \left(1 - \frac{c_1}{\left(\frac{\varepsilon}{n\,\rho_n}\right)^{1/r}}\right)^n = \\ = 1 - \left(1 - \frac{c_1\, n^{\frac{r-s+1}{r}} \left(n^s \rho_n\right)^{\frac{1}{r}}}{\varepsilon^{1/r}}\right)^n \sim \frac{c_1}{\varepsilon^{1/r}}\, n^{\frac{r-s+1}{r}} \left(n^s \rho_n\right)^{\frac{1}{r}} \sim n^{\frac{r-s+1}{r}}, \quad n \to \infty.\end{aligned} \tag{5.1.1}$$

■

Enfraquecendo a anterior condição sobre a velocidade de convergência para zero do raio espectral ρ_n conseguimos obter ainda a convergência em probabilidade do EMQ.

Corolário 5.1.1 *Se $r > 0$ e no modelo de regressão linear* (5.0.1),

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.*

(b) $\overline{F}_{\mathrm{e}^2}(x) = O\big(x^{-1/r}\big)$, $x \to \infty$.

(c) $\exists s > r+1\colon\ \rho_n = O\big(1/n^s\big)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ é fracamente consistente.

Demonstração. Da expressão obtida em (5.1.1) tem-se

$$\lim_{n\to\infty} \mathbb{P}\big\{n\rho_n \mathrm{M}_n > \varepsilon\big\} \leqslant c_1 \left(\frac{c_2}{\varepsilon}\right)^{1/r} \lim_{n\to\infty} n^{\frac{r-s+1}{r}} = 0,$$

dado que estamos a assumir que $r > 0$ e $s > r+1$. ■

Exemplo 5.1.1

Como exemplo de aplicação, podemos pensar no caso em que os erros têm distribuição de Cauchy com parâmetros $(0, \pi)$,

$$f_{\mathrm{e}}(x) = \frac{1}{\pi^2 + x^2}, \quad x \in \mathbb{R}.$$

A densidade do quadrado dos erros será,

$$f_{\mathrm{e}^2}(x) = \frac{f_{\mathrm{e}}\left(\sqrt{x}\right) + f_{\mathrm{e}}\left(-\sqrt{x}\right)}{2\sqrt{x}}, \quad x > 0$$

tendo-se $f_{\mathrm{e}^2}(x) \sim x^{-\frac{3}{2}}$, $x \to \infty$ e $\overline{F}_{\mathrm{e}^2}(x) \sim 2x^{-\frac{1}{2}}$, $x \to \infty$ donde bastará tomar $r = 2$ e escolher $\rho_n = O\big(1/n^s\big)$, $n \to \infty$ com $s > 3$ e $s > 5$ para se obter, respectivamente, a consistência fraca e forte do EMQ. Convém destacar que, nesta situação, a v.a. e^2 não possui momento absoluto de ordem 1.

Observação 5.1.1 A condição sobre o raio espectral em [43] é $(\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1} = O\big(n^{-(2-\alpha)/\alpha}\big)$, $n \to \infty$ onde as componentes do vector dos erros verificam $\mathbb{E}\,|\mathrm{e}_1|^{\alpha} < \infty$ com $\alpha \geqslant 1$. É de fácil constatação que as hipóteses $\rho_n = O\big(1/n^s\big)$, $n \to \infty$ com $s > 2r+1$ e $\overline{F}_{\mathrm{e}^2}(x) = O(x^{-1/r})$, $x \to \infty$ são claramente mais fortes do que as admitidas em [43] mas existe a óbvia vantagem deste resultado poder ser aplicado à situação em que $0 < \alpha < 1$.

É sabido que, na classe de distribuições que possuem suporte não limitado à direita encontram-se, em particular, todas as distribuições que estão no domínio de atracção maximal da distribuição de Fréchet.

Teorema 5.1.2 *Se no modelo de regressão linear* (5.0.1),

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.*

(b) $F_{\mathrm{e}^2} \in \mathcal{D}(\Phi_{1/r})$ $(r > 0)$ *satisfazendo* $\overline{F}_{\mathrm{e}^2}(x) = x^{-1/r}\ell(x)$ *e*

$$\rho_n = \begin{cases} o\big(n^{-1}\big) & \textit{se} \quad r < 1 \ \textit{ou} \ \left(r = 1 \ \textit{e} \ \displaystyle\sum_{n=1}^{\infty} \frac{\ell(n)}{n} < \infty\right) \\ O\left(\dfrac{1}{b_n^r\, \ell_1(n)}\right) & \textit{se} \quad \textit{caso contrário} \end{cases}, \qquad n \to \infty$$

com $b_n > 0$ *uma qualquer sucessão de reais tal que* $\dfrac{b_n}{n}$ *é crescente e* $\displaystyle\sum_{n=1}^{\infty} \frac{1}{b_n} < \infty$[1] *sendo* $\ell_1(x)$ *a função de variação lenta definida por* $\ell_1(x) = x^{-r} \cdot F^{\leftarrow}_{e^2}\big(1 - x^{-1}\big) \sim \left[\left(\dfrac{1}{\ell}\right)^r\right]^{\#}(x^r)$*,* $x \to \infty$*.*

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Se $r < 1$ ou $r = 1$ e $\displaystyle\sum_{n=1}^{\infty} \frac{\ell(n)}{n} < \infty$[2] então $\mathbb{E}\big(e^2\big) < \infty$ ficando a tese estabelecida através da clássica lei forte de Kolmogorov (conf. Observação 1.1.7). Se caso contrário tem-se sempre $\mathbb{E}\big(e^2\big) = \infty$ e o resultado segue pelo Teorema 1.1.18 já que,

$$\overline{F}_{e^2}\big(b_n^r\,\ell_1(n)\big) \sim \frac{\big(n^{-r}\,b_n^r\big)^{-1/r}}{n} = \frac{1}{b_n}, \qquad n \to \infty. \qquad \blacksquare$$

Corolário 5.1.2 *Se no modelo de regressão linear* (5.0.1)*,*

(a) $e, e_1, e_2, \ldots$ *são i.i.d. tais que* $F_e(-x) = O\left(\overline{F}_e(x)\right)$*,* $x \to \infty$*.*

(b) $F_e \in \mathcal{D}(\Phi_{1/r})$ $(r > 0)$ *satisfazendo* $\overline{F}_e(x) = x^{-1/r}\ell(x)$ *e*

$$\rho_n = \begin{cases} o\big(n^{-1}\big) & \textit{se} \quad r < 1 \ \textit{ou} \ \left(r = 1 \ \textit{e} \ \displaystyle\sum_{n=1}^{\infty} \frac{\ell\left(\sqrt{n}\right)}{n} < \infty\right) \\ O\left(\dfrac{1}{b_n^{2r}\,\ell_1(n)}\right) & \textit{se} \quad \textit{caso contrário} \end{cases}, \qquad n \to \infty$$

com $b_n > 0$ *uma qualquer sucessão de reais tal que* $\dfrac{b_n}{n}$ *é crescente e* $\displaystyle\sum_{n=1}^{\infty} \frac{1}{b_n} < \infty$ *sendo* $\ell_1(x)$ *a função de variação lenta definida por* $\ell_1(x) \sim \left\{\left[\left(\dfrac{1}{\ell}\right)^r\right]^{\#}(x^r)\right\}^2$*,* $x \to \infty$*.*

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Com efeito, nesta situação tem-se,

$$\overline{F}_{e^2}(x) = \overline{F}_e(\sqrt{x}) + F_e(-\sqrt{x}) = x^{-\frac{1}{2r}}\,\widehat{\ell}(x)\big(1 + O(1)\big), \quad x \to \infty \tag{5.1.2}$$

com $\widehat{\ell}(x) = \ell(\sqrt{x})$ e a conclusão segue do Teorema 5.1.1. $\blacksquare$

O próximo resultado dá-nos a consistência em média quadrática do EMQ.

Teorema 5.1.3 *Se no modelo de regressão linear* (5.0.1)*,*

(a) $e, e_1, e_2, \ldots$ *são identicamente distribuídos.*

(b) $F_{e^2} \in \mathcal{D}(\Phi_{1/r})$*,* $0 < r < 1$*.*

[1]Note-se que, em particular, podemos tomar $b_n = n\log n \cdot \log\log(n) \ldots (\log_m(n))^p$ $(p > 1)$ onde $\log_m n = \log(\log(\ldots\log n))$ para algum $m \in \mathbb{N}$.

[2]Observe-se que $\mathbb{E}\big(e^2\big) < \infty \iff \displaystyle\sum_{n=1}^{\infty} \frac{\ell(n)}{n} < \infty$.

(c) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. Visto que $0 < r < 1$ então $\mathbb{E}\left(\mathrm{e}^2\right) < \infty$ donde a estimativa (5.0.2) implica,

$$\mathbb{E}\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2\right) \leqslant \rho_n\, \mathbb{E}\left(\|\mathbf{e}_n\|^2\right) = n\,\rho_n \cdot \mathbb{E}\left(\mathrm{e}^2\right) \tag{5.1.3}$$

e a hipótese (c) estabelece a tese. ∎

Corolário 5.1.3 *Se no modelo de regressão linear* (5.0.1)*,*

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são identicamente distribuídos tais que* $F_{\mathrm{e}}(-x) = O\left(\overline{F}_{\mathrm{e}}(x)\right)$, $x \to \infty$.

(b) $F_{\mathrm{e}} \in \mathcal{D}(\Phi_{1/r})$, $0 < r < \frac{1}{2}$.

(c) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. Considerando $\overline{F}_{\mathrm{e}}(x) = x^{-\frac{1}{r}}\,\ell(x)$, $0 < r < \frac{1}{2}$ para alguma função de variação lenta ℓ então das hipóteses assumidas obtém-se $\overline{F}_{\mathrm{e}^2}(x) = x^{-\frac{1}{2r}}\,\widehat{\ell}(x)\big(1 + O(1)\big)$, $x \to \infty$ com $\widehat{\ell}(x) = \ell(\sqrt{x})$ e o Teorema 5.1.3 conclui a demonstração. ∎

Observação 5.1.2 Observemos que se $0 < s \leqslant 2$ e no modelo de regressão linear (5.0.1) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ são identicamente distribuídas, $F_{\mathrm{e}^2} \in \mathcal{D}(\Phi_{1/r})$ com $0 < rs < 2$ e $\rho_n = o\left(n^{-\frac{2}{s}}\right)$, $n \to \infty$ então $\widetilde{\boldsymbol{\beta}}$ é consistente em momento de ordem s.

Observação 5.1.3 Variáveis aleatórias com distribuição de Pareto, Cauchy, Cauchy generalizada, Burr, log-gama (ver Apêndice) ou estáveis com índice de estabilidade $\alpha < 2$ (ver Secção 1.4) estão no domínio de atracção maximal da distribuição de Fréchet.

5.2 O domínio de atracção maximal da distribuição Weibull

Avançamos, agora, para a análise das distribuições que pertencem ao domínio de atracção maximal da distribuição de Weibull.

Teorema 5.2.1 *Se no modelo de regressão linear* (5.0.1)*,*

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.*

(b) $F_{\mathrm{e}^2} \in \mathcal{D}(\Psi_r)$, $r > 0$.

(c) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Sendo $x_{\sup}$ o extremo direito do suporte de F_{e^2} tem-se para todo o $x < x_{\sup}$,

$$\mathbb{P}\big(\mathrm{M}_n \leqslant x\big) = F^n(x) \longrightarrow 0, \quad n \to \infty.$$

Quando $x \geqslant x_{\sup}$ vem $\mathbb{P}\big(\mathrm{M}_n \leqslant x\big) = F^n(x) = 1$ donde $\mathrm{M}_n \xrightarrow{\mathbb{P}} x_{\sup}$. Consequentemente, a sucessão $\mathrm{M}_1, \mathrm{M}_2, \ldots$ convergirá quase certamente uma vez que é não-decrescente em n,

$$\mathrm{M}_n \xrightarrow{\text{q.c}} x_{\sup} \tag{5.2.1}$$

(ver [57]) e a tese fica estabelecida através da estimativa (5.0.2). ∎

Corolário 5.2.1 *Se no modelo de regressão linear* (5.0.1),

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.*

(b) *O extremo esquerdo e direito do suporte de* F_{e}, *respectivamente,* $x_{\inf}$ *e* $x_{\sup}$ *satisfazem* $|x_{\inf}| \leqslant |x_{\sup}|$ *e* $F_{\mathrm{e}}(-x) = O\big(\overline{F}_{\mathrm{e}}(x)\big)$, $x \to x_{\sup}$.

(c) $F_{\mathrm{e}} \in \mathcal{D}(\Psi_r)$, $r > 0$.

(d) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Com efeito,

$$\overline{F}_{\mathrm{e}^2}(x) = \overline{F}_{\mathrm{e}}\big(\sqrt{x}\big) + F_{\mathrm{e}}\big(-\sqrt{x}\big) = \overline{F}_{\mathrm{e}}\big(\sqrt{x}\big)\big(1 + O(1)\big) = \overline{G}\big(\sqrt{x}\big)\big(1 + O(1)\big), \quad x \to x_{\sup}^2$$

onde G é a distribuição definida por $G(u) = F_{\mathrm{e}}(u)$ se $u \geqslant 0$ e $G(u) = 0$ se $u < 0$ (note-se que $F_{\mathrm{e}^2}(x) = 0$ quando $x < 0$). Então $F_{\mathrm{e}^2}(x)$ e $G\big(\sqrt{x}\big)$ são de cauda equivalente (conf. Definição 1.1.9) e portanto $F_{\mathrm{e}^2}(x)$ pertence ao domínio de atracção maximal da distribuição de Weibull $\Psi_{r/2}$ $(r > 0)$. A tese é agora uma consequência do Teorema 5.2.1. ∎

De seguida, anunciamos a consistência em média quadrática do EMQ quando a distribuição do quadrado dos erros pertence ao domínio de atracção maximal da distribuição de Weibull.

Teorema 5.2.2 *Se no modelo de regressão linear* (5.0.1),

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são identicamente distribuídos.*

(b) $F_{\mathrm{e}^2} \in \mathcal{D}(\Psi_r)$, $r > 0$.

(c) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. Voltando a denotar o extremo direito do suporte de F_{e^2} por $x_{\sup}$ conclui-se, sem dificuldade, que $\mathbb{E}(\mathrm{M}_n) \longrightarrow x_{\sup}$ quando $n \to \infty$ uma vez que a sucessão M_n é positiva, não-decrescente e verifica (5.2.1) (conf. Corolário 2 da Secção 4.2 de [18]). A tese fica estabelecida através da estimativa,

$$\mathbb{E}\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2\right) \leqslant n\,\rho_n \cdot \mathbb{E}(\mathrm{M}_n). \tag{5.2.2}$$

■

Corolário 5.2.2 *Se no modelo de regressão linear* (5.0.1),

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são identicamente distribuídos.*

(b) *O extremo esquerdo e direito do suporte de* F_{e}*, respectivamente,* $x_{\inf}$ *e* $x_{\sup}$ *satisfazem* $|x_{\inf}| \leqslant |x_{\sup}|$ *e* $F_{\mathrm{e}}(-x) = O\big(\overline{F}_{\mathrm{e}}(x)\big)$, $x \to x_{\sup}$.

(c) $F_{\mathrm{e}} \in \mathcal{D}(\Psi_r)$, $r > 0$.

(d) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. Das hipóteses assumidas vem $\overline{F}_{e^2}(x) = \overline{G}\big(\sqrt{x}\big)\big(1 + O(1)\big)$, $x \to x^2_{\sup}$ onde G é a distribuição definida por $G(u) = F_e(u)$ se $u \geqslant 0$ e $G(u) = 0$ se $u < 0$. O Teorema 5.2.2 conclui a demonstração já que $F_{e^2}(x)$ e $G\big(\sqrt{x}\big)$ são de cauda equivalente. ■

Observação 5.2.1 Na hipótese (b) dos Corolários 5.2.1 e 5.2.2 se $|x_{\inf}| < |x_{\sup}|$ então a condição $F_{\mathrm{e}}(-x) = O\big(\overline{F}_{\mathrm{e}}(x)\big)$, $x \to x_{\sup}$ é trivialmente cumprida uma vez que $F_{\mathrm{e}}(-x) = 0$, $\forall x \in \big]\,|x_{\inf}|, x_{\sup}\big[$.

Observação 5.2.2 Estes resultados podem ser aplicados a variáveis aleatórias com distribuição uniforme, beta ou ainda a caudas de probabilidade com comportamento de potência x^r $(r > 0)$ no extremo direito do suporte $x_{\sup}$ limitado:

$$\overline{F}(x) = K\left(x_{\sup} - x\right)^r, \qquad x_{\sup} - K^{-1/r} \leqslant x \leqslant x_{\sup}$$

para $K, r > 0$.

5.3 O domínio de atracção maximal da distribuição Gumbel

Finalizamos este capítulo com o estudo da convergência forte e em média quadrática do EMQ quando a função de distribuição de e^2 pertence à terceira grande classe de distribuições: o domínio de atracção maximal da distribuição de Gumbel.

Teorema 5.3.1 *Se no modelo de regressão linear* (5.0.1),

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.*

(b) $F_{e^2} \in \mathcal{D}(\Lambda)$.

(c) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Seja $x_{\sup}$ o extremo direito do suporte de F_{e^2}. Se $x_{\sup} = \infty$ tem-se $\mathbb{E}\left(e^2\right) < \infty$ (conf. Proposição 1.5.2) e a tese fica estabelecida pela lei forte de Kolmogorov (conf. Observação 1.1.7). Se $x_{\sup} < \infty$ a tese resulta automaticamente da convergência (5.2.1). ∎

Corolário 5.3.1 *Se no modelo de regressão linear* (5.0.1),

(a) $e, e_1, e_2, \ldots$ *são i.i.d.*

(b) *O extremo direito do suporte de* F_e *for ilimitado e* $F_e(-x) = O\left(\overline{F}_e(x)\right)$, $x \to \infty$.

(c) $F_e \in \mathcal{D}(\Lambda)$.

(d) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é fortemente consistente.*

Demonstração. Com efeito, quando $x \to \infty$ tem-se,

$$\overline{F}_{e^2}(x) = \overline{F}_e\left(\sqrt{x}\right) + F_e\left(-\sqrt{x}\right) = h\left(\sqrt{x}\right) \exp\left\{-\int_{x_0^2}^{x} \frac{g\left(\sqrt{t}\right)}{2\sqrt{t}\,\alpha\left(\sqrt{t}\right)}\,dt\right\}\left(1 + O(1)\right)$$

e $\lim_{x\to\infty} h\left(\sqrt{x}\right) = c_0 > 0$, $\lim_{x\to\infty} g\left(\sqrt{x}\right) = 1$. A função $\widehat{\alpha}(x) = 2\sqrt{x}\,\alpha\left(\sqrt{x}\right)$ é positiva, absolutamente contínua em $]0, \infty[$ e verifica,

$$\widehat{\alpha}'(x) = \frac{\alpha(\sqrt{x})}{\sqrt{x}} + \alpha'(\sqrt{x}) \longrightarrow 0, \quad x \to \infty$$

já que $\alpha'(u) \longrightarrow 0$ quando $u \to \infty$ permite escrever,

$$\lim_{u\to\infty} \frac{\alpha(u)}{u} = \lim_{u\to\infty} \frac{1}{u} \int_{u_0}^{u} \alpha'(t)\,dt = 0$$

isto é, $\alpha(u) = o(u)$, $u \to \infty$. A tese é agora uma consequência do Teorema 5.3.1. ∎

Corolário 5.3.2 *Se no modelo de regressão linear* (5.0.1),

(a) $e, e_1, e_2, \ldots$ *são i.i.d.*

(b) *O extremo esquerdo e direito do suporte de* F_e*, respectivamente,* $x_{\inf}$ *e* $x_{\sup}$ *satisfazem* $|x_{\inf}| \leqslant |x_{\sup}| < \infty$ *e* $F_e(-x) = O\left(\overline{F}_e(x)\right)$, $x \to x_{\sup}$.

(c) F_e *é uma função de von Mises.*

(d) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ é fortemente consistente.

Demonstração. De facto,

$$\overline{F}_{e^2}(x) = \overline{F}_{e}\big(\sqrt{x}\big) + F_{e}\big(-\sqrt{x}\big) = c\exp\left\{-\int_{x_0^2}^{x}\frac{1}{\widehat{\alpha}(t)}\,dt\right\}\big(1+O(1)\big), \quad x \to x_{\sup}^2$$

em que $\widehat{\alpha}(x) = 2\sqrt{x}\,\alpha\big(\sqrt{x}\big)$ é positiva, absolutamente contínua em $\left]0, x_{\sup}^2\right[$ e satisfaz,

$$\widehat{\alpha}'(x) = \frac{\alpha(\sqrt{x})}{\sqrt{x}} + \alpha'(\sqrt{x}) \longrightarrow 0, \quad x \to x_{\sup}^2$$

uma vez que $\alpha(u) = o(x_{\sup} - u) = o(1)$, $u \to x_{\sup}$ (ver alínea (b) do Lema 1.2 de [78]). A conclusão da demonstração segue mais uma vez pelo Teorema 5.3.1. ∎

A consistência em média quadrática do EMQ é apresentada no próximo resultado.

Teorema 5.3.2 *Se no modelo de regressão linear* (5.0.1)*,*

(a) $e, e_1, e_2, \ldots$ *são identicamente distribuídos.*

(b) $F_{e^2} \in \mathcal{D}(\Lambda)$.

(c) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ é consistente em média quadrática.

Demonstração. Mencionando novamente o extremo direito do suporte de F_{e^2} por $x_{\sup}$ temos $\mathbb{E}\big(e^2\big) < \infty$ se $x_{\sup} = \infty$ e a tese fica estabelecida através de (5.1.3). Se $x_{\sup} < \infty$ a conclusão obtém-se raciocinando como na demonstração do Teorema 5.2.2. ∎

Corolário 5.3.3 *Se no modelo de regressão linear* (5.0.1)*,*

(a) $e, e_1, e_2, \ldots$ *são identicamente distribuídos.*

(b) *O extremo direito do suporte de F_e for ilimitado e $F_e(-x) = O\left(\overline{F}_e(x)\right)$, $x \to \infty$.*

(c) $F_e \in \mathcal{D}(\Lambda)$.

(d) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ é consistente em média quadrática.

Demonstração. Na demonstração do Corolário 5.3.1 vimos que as hipóteses assumidas conduzem-nos a,

$$\overline{F}_{e^2}(x) = \widehat{h}(x)\exp\left\{-\int_{x_0^2}^{x}\frac{\widehat{g}(t)}{\widehat{\alpha}(t)}\,dt\right\}\big(1+O(1)\big), \quad x \to \infty$$

onde $\widehat{h}(x) = h\big(\sqrt{x}\big)$, $\widehat{g}(x) = g\big(\sqrt{x}\big)$ verificam $\lim_{x\to\infty}\widehat{h}(x) = c_0 > 0$, $\lim_{x\to\infty}\widehat{g}(x) = 1$ e $\widehat{\alpha}(x) = 2\sqrt{x}\,\alpha\big(\sqrt{x}\big)$ é positiva, absolutamente contínua em $\left]0, \infty\right[$ e satisfaz $\lim_{x\to\infty}\widehat{\alpha}'(x) = 0$. A tese é agora uma consequência do Teorema 5.3.2. ∎

Corolário 5.3.4 *Se no modelo de regressão linear* (5.0.1),

(a) $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são identicamente distribuídos.*

(b) *O extremo esquerdo e direito do suporte de* F_{e}*, respectivamente,* $x_{\inf}$ *e* $x_{\sup}$ *satisfazem* $|x_{\inf}| \leqslant |x_{\sup}| < \infty$ *e* $F_{\mathrm{e}}(-x) = O\big(\overline{F}_{\mathrm{e}}(x)\big)$, $x \to x_{\sup}$.

(c) F_{e} *é uma função de von Mises.*

(d) $\rho_n = o\left(n^{-1}\right)$, $n \to \infty$.

então $\widetilde{\boldsymbol{\beta}}$ *é consistente em média quadrática.*

Demonstração. Já vimos na demonstração do Corolário 5.3.2 que,

$$\overline{F}_{\mathrm{e}^2}(x) = c \exp\left\{ -\int_{x_0^2}^{x} \frac{1}{\widehat{\alpha}(t)}\, dt \right\} \big(1 + O(1)\big), \quad x \to x_{\sup}^2$$

onde $\widehat{\alpha}(x) = 2\sqrt{x}\, \alpha\big(\sqrt{x}\big)$ é positiva, absolutamente contínua em $\left]0, x_{\sup}^2\right[$ e satisfaz $\lim\limits_{x \to x_{\sup}^2} \widehat{\alpha}'(x) = 0$. A conclusão da prova segue então pelo Teorema 5.3.2. ∎

Observação 5.3.1 Na hipótese (b) dos Corolários 5.3.2 e 5.3.4 se $|x_{\inf}| < |x_{\sup}| < \infty$ então a condição $F_{\mathrm{e}}(-x) = O\big(\overline{F}_{\mathrm{e}}(x)\big)$, $x \to x_{\sup}$ é imediatamente verificada pois $F_{\mathrm{e}}(-x) = 0$, $\forall x \in \left]\, |x_{\inf}|, x_{\sup}\right[$.

Observação 5.3.2 Nesta linha de resultados enquadra-se qualquer distribuição que tenha comportamento exponencial no extremo direito do suporte limitado como,

$$\overline{F}(x) = K \exp\left(\frac{C}{x - x_{\sup}}\right), \qquad x < x_{\sup}$$

para $K, C > 0$ ou no caso em que o extremo direito do suporte é ilimitado, distribuições com o seguinte comportamento de cauda: $\overline{F}(x) = \exp\left(1 - e^x\right)$, $x > 0$. As situações em que o quadrado dos erros têm distribuição de Weibull-Gnedenko, exponencial, normal, log-normal, Gama, Benktander tipo I, Benktander tipo II ou de Erlang estão também incluídas neste domínio de atracção maximal (ver Apêndice).

5.4 Exemplos de matrizes de modelo

Uma questão natural ligada aos resultados desenvolvidos anteriormente é a existência de matrizes de modelo com velocidade de convergência para zero arbitrária quando o número de observações cresce infinitamente. Na realidade, é o caso em que o peso nas observações mais recentes cresce polinomialmente.

Teorema 5.4.1 *Dado um inteiro positivo* p *existe uma matriz de modelo* $\boldsymbol{X}_n(p)$ *tal que*

$$\lim_{n \to \infty} n^{2p+1} \cdot \rho\left(\left(\boldsymbol{X}_n^T(p)\, \boldsymbol{X}_n(p)\right)^{-1}\right) = (2p+1)\,(2p+2)^2.$$

Demonstração. Consideremos a seguinte matriz de modelo

$$\boldsymbol{X}_n(p) = \begin{bmatrix} 1^p & 1^{p+1} \\ 2^p & 2^{p+1} \\ \vdots & \vdots \\ n^p & n^{p+1} \end{bmatrix}.$$

Então,

$$\boldsymbol{X}_n^T(p)\,\boldsymbol{X}_n(p) = \begin{bmatrix} S_{2p}(n) & S_{2p+1}(n) \\ S_{2p+1}(n) & S_{2p+2}(n) \end{bmatrix},$$

onde $S_p(n) = 1^p + 2^p + \ldots + n^p$. Os valores próprios da matriz $\left(\boldsymbol{X}_n^T(p)\,\boldsymbol{X}_n(p)\right)^{-1}$ são não-negativos e o seu raio espectral é

$$\rho_n = \frac{S_{2p}(n) + S_{2p+2}(n) + \sqrt{\left(S_{2p}(n) - S_{2p+2}(n)\right)^2 + 4\left(S_{2p+1}(n)\right)^2}}{2\left(S_{2p}(n)\,S_{2p+2}(n) - \left(S_{2p+1}(n)\right)^2\right)}.$$

Da fórmula de Faulhaber (ver [20]) sabemos que

$$1^p + 2^p + \ldots + n^p = \sum_{j=1}^{p+1} b_{pj}\, n^j,$$

com $b_{pj} = \dfrac{(-1)^{\delta_{jp}}\, \beta_{p-j+1}\, p!}{j!\,(p-j+1)!}$ e β_i $(i = 0, 1, \ldots, p)$ os números de Bernoulli (ver Secção 1.3). Deste modo,

$$S_{2p}(n) + S_{2p+2}(n) = \sum_{j=1}^{2p+3} u_j\, n^j \quad \text{com} \quad u_{2p+3} = \frac{1}{2p+3},$$

$$\left(S_{2p}(n) - S_{2p+2}(n)\right)^2 + 4\left(S_{2p+1}(n)\right)^2 = \sum_{j=1}^{4p+6} v_j\, n^j \ \text{ com } \ v_{4p+6} = \frac{1}{(2p+3)^2}$$

e

$$2\left(S_{2p}(n)\,S_{2p+2}(n) - \left(S_{2p+1}(n)\right)^2\right) = \sum_{j=1}^{4p+4} z_j\, n^j \ \text{ com } \ z_{4p+4} = \frac{2}{(2p+1)(2p+2)^2(2p+3)}.$$

Através destas expressões obtém-se

$$\lim_{n\to+\infty} n^{2p+1} \cdot \rho_n = \frac{\dfrac{2}{2p+3}}{\dfrac{2}{(2p+1)\,(2p+2)^2\,(2p+3)}} = (2p+1)\,(2p+2)^2,$$

o que estabelece a tese. ∎

Observação 5.4.1 No caso em estudo, podemos concluir que dado um modelo possuindo (eventualmente) erros de elevada potência no índice de cauda é sempre possível encontrar uma matriz de modelo tal que o estimador dos mínimos quadrados seja fortemente consistente.

Capítulo 6

Estimadores não-lineares

Seguindo o exemplo apresentado na página 139 de [21], suponhamos uma experiência que consiste na emissão de um sinal de coordenadas a_i, b_i, c_i transformado em sinal numérico $f(a_i, b_i, c_i)$ por um determinado aparelho e recebido sobre uma linha de ruído. A recepção do sinal pode ser traduzida pelo modelo,

$$\mathrm{Y}_i = f(a_i, b_i, c_i) + \mathrm{e}_i \tag{6.0.1}$$

onde Y_i representa a i-ésima recepção, $f(a_i, b_i, c_i)$ a transformação do i-ésimo sinal e e_i o i-ésimo ruído (variável aleatória com média nula e distribuição independente de i). Efectuadas n emissões com diferentes doses de sinal o investigador terá como objectivo achar a dose optimal de sinal. O modelo (6.0.1) pode ser enquadrado no seguinte modelo para as observações,

$$\mathrm{Y}_i = \mu_i + \mathrm{e}_i, \quad i = 1, \ldots, n$$

ou, em notação matricial,

$$\mathbf{Y}_n = \boldsymbol{\mu}_n + \mathbf{e}_n \tag{6.0.2}$$

com

$$\mathbf{Y}_n = \begin{bmatrix} \mathrm{Y}_1 \\ \vdots \\ \mathrm{Y}_n \end{bmatrix}, \quad \boldsymbol{\mu}_n = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \quad \text{e} \quad \mathbf{e}_n = \begin{bmatrix} \mathrm{e}_1 \\ \vdots \\ \mathrm{e}_n \end{bmatrix}.$$

À semelhança dos capítulos anteriores, voltamos a destacar o índice n nos vectores envolvidos por uma questão de passagens ao limite em n. De um modo geral, podemos idealizar um estimador da forma $\boldsymbol{\Theta}_n = \boldsymbol{g}_n(\mathbf{Y}_n)$ onde $\boldsymbol{g}_n$ é uma aplicação regular (num sentido que adiante tornaremos mais preciso) definida em $\mathbb{R}^n$ e tomando valores em $\mathbb{R}^\kappa$ com $\kappa \leqslant n$ fixo.

Exemplo 6.0.1

Um exemplo conhecido acontece quando $g_n \colon \mathbb{R}^n \longrightarrow \mathbb{R}$ está definida por $g_n(x_1, \ldots, x_n) = \frac{x_1 + \ldots + x_n}{n}$ e o estimador toma a forma,

$$\Theta_n = \frac{\mu_1 + \ldots + \mu_n}{n} + \frac{\mathrm{e}_1 + \ldots + \mathrm{e}_n}{n}.$$

Se a sucessão de números reais $\{\mu_n\}$ converge para $\mu_\infty \in \mathbb{R}$ e a sucessão da v.a.'s $\{\mathrm{e}_i,\, n \geqslant 1\}$ for i.i.d. tal que $\mathbb{E}\,|\mathrm{e}_i| < \infty$, $\forall i$ então $\lim\limits_{n\to\infty} \frac{\mathrm{e}_1+\ldots+\mathrm{e}_n}{n} = \mathbb{E}(\mathrm{e})$ q.c. (conf. Observação 1.1.7) e $\lim\limits_{n\to\infty} \frac{\mu_1+\ldots+\mu_n}{n} = \mu_\infty$ (ver Proposição 1.3.1) donde a sucessão dos estimadores $\{\Theta_n,\, n \geqslant 1\}$ converge fortemente para $\mu + \mathbb{E}(\mathrm{e})$.

Em suma, nas próximas secções analisaremos funções $\boldsymbol{g}_n : \mathbb{R}^n \longrightarrow \mathbb{R}^\kappa$ não necessariamente lineares e apresentaremos condições pouco restritivas que conduzirão à convergência quase certa da sucessão de estimadores $\{\boldsymbol{\Theta}_n,\, n \geqslant 1\}$.

6.1 Variáveis aleatórias com decaimento exponencial

Algumas distribuições de probabilidade como a distribuição normal, χ^2, gama, Weibull, entre outras, exibem um comportamento assimptótico notável que clarificamos em seguida.

Definição 6.1.1 Uma variável aleatória X tem **decaimento exponencial no infinito** (**DEI**) de parâmetros $\alpha, \beta > 0$ e $\tau \in \mathbb{R}$ se $|\mathrm{X}|$ tiver função densidade de probabilidade,

$$f_{|\mathrm{X}|}(x) \asymp x^\tau e^{-\beta x^\alpha}, \quad x \to \infty. \tag{6.1.1}$$

Se X tiver DEI com parâmetros $\alpha, \beta > 0$ e $\tau \in \mathbb{R}$ escrevemos $\mathrm{X} \sim \mathrm{DEI}(\alpha, \beta, \tau)$.

A proposição seguinte justifica a denominação dada às variáveis aleatórias introduzidas na Definição 6.1.1.

Proposição 6.1.1 *Se* $\mathrm{X} \sim \mathrm{DEI}(\alpha, \beta, \tau)$ *então* $\mathbb{P}\{\,|\mathrm{X}| > u\} = o\left(u^{\tau+1} e^{-\beta u^\alpha}\right)$, $u \to \infty$.

Demonstração. Comecemos por observar que a função $h(x) = e^{-\beta x^\alpha}$, $x > 0$ é não-crescente e de variação rápida no infinito já que satisfaz,

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = \lim_{x\to\infty} e^{-\beta(t-1)x^\alpha} = \begin{cases} \infty & \text{se} \quad 0 < t < 1 \\ 1 & \text{se} \quad t = 1 \\ 0 & \text{se} \quad t > 1 \end{cases}.$$

Então,

$$\mathbb{P}\{\,|\mathrm{X}| > u\} = \int_u^\infty f_{|\mathrm{X}|}(x)\,dx \leqslant c_2 \int_u^\infty x^\tau e^{-\beta x^\alpha}\,dx = o\left(u^{\tau+1} e^{-\beta u^\alpha}\right), \quad u \to \infty \qquad (c_2 > 0)$$

pois o Teorema 1.3.4 assegura que $\int_u^\infty x^\tau e^{-\beta x^\alpha}\,dx = o\left(u^{\tau+1} e^{-\beta u^\alpha}\right)$, $u \to \infty$. ∎

Observação 6.1.1 Note-se que uma variável aleatória $\mathrm{X} \sim \mathrm{DEI}(\alpha, \beta, \tau)$ tem momentos absolutos de todas as ordens. Com efeito, para qualquer $p > 0$ tem-se $u^{p-1}\mathbb{P}\{\,|X| > u\} = o\left(u^{\tau+p} e^{-\beta u^\alpha}\right)$, $u \to \infty$ e $\int_0^\infty u^{\tau+p} e^{-\beta u^\alpha}\,du = \frac{1}{\alpha\, \beta^{\frac{\tau+p+1}{\alpha}}}\, \Gamma\left(\frac{\tau+p+1}{\alpha}\right)$ o que implica,

$$\mathbb{E}\,|X|^p = p \int_0^\infty u^{p-1} \mathbb{P}\{\,|X| > u\}\,du < \infty.$$

Exemplo 6.1.1
Se $X \geqslant 0$ e $X \sim \text{DEI}(\alpha,\beta,\tau)$ então $\sqrt{X} \sim \text{DEI}\,(2\alpha,\beta,2\tau+1)$. Se $X \sim \text{DEI}(\alpha,\beta,\tau)$ então $X^2 \sim \text{DEI}\left(\frac{\alpha}{2},\beta,\frac{\tau-1}{2}\right)$.

Proposição 6.1.2 *Seja* $X, X_1, X_2, \ldots$ *uma sucessão de variáveis aleatórias identicamente distribuídas tal que* $X \sim \text{DEI}(\alpha,\beta,\tau)$ *e* $\{a_n\}$ *uma sucessão de reais positiva satisfazendo* $\lim\limits_{n\to\infty} a_n = \infty$ *e* $\sum_{n=1}^{\infty} e^{-\delta a_n^{\alpha}} < \infty$ *para qualquer* $\delta > 0$. *Então*

$$\frac{X_n}{a_n} \xrightarrow{\text{q.c.}} 0.$$

Demonstração. De acordo com a Proposição 6.1.1 tem-se para todo o $\varepsilon > 0$,

$$\mathbb{P}\left\{\frac{|X_n|}{a_n} > \varepsilon\right\} = o\left(\left(a_n\,\varepsilon\right)^{\tau+1} e^{-\beta(\varepsilon a_n)^{\alpha}}\right), \quad n \to \infty.$$

Escolhendo $\delta < \beta\varepsilon^{\alpha}$ em $\sum_{n=1}^{\infty} e^{-\delta a_n^{\alpha}} < \infty$ garantimos

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{\frac{|X_n|}{a_n} > \varepsilon\right\} < \infty$$

o que implica $\dfrac{X_n}{a_n} \xrightarrow{\text{q.c.}} 0$ (ver Teorema 1.1.21). ■

6.2 A convergência forte

Suponhamos que, para todo o $n \in \mathbb{N}$, se tem que $\boldsymbol{g}_n \in C^1(\mathbb{R}^n;\mathbb{R}^{\kappa})$ e representemos por $\mathbf{D}\boldsymbol{g}_n$ a aplicação diferencial de $\boldsymbol{g}_n$. Em concordância com o modelo que supusemos verificar-se vem,

$$\boldsymbol{\Theta}_n = \boldsymbol{g}_n(\mathbf{Y}_n) = \boldsymbol{g}_n(\boldsymbol{\mu} + \mathbf{e}_n) = \boldsymbol{g}_n(\boldsymbol{\mu}) + \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\mathbf{e}_n \tag{6.2.1}$$

em que $\boldsymbol{\lambda}_n$ é um ponto aleatório do conjunto $\mathbb{L} = \left\{t\mathbf{Y}_n + (1-t)\boldsymbol{\mu}_n \colon t \in]0,1[\right\}$. O lado direito da fórmula resulta da aplicação do teorema do valor médio de Lagrange à função $h(t) = \boldsymbol{g}_n\big(t\mathbf{Y}_n + (1-t)\boldsymbol{\mu}_n\big)$.

Supondo que existe uma sucessão numérica $\{a_n\}$ positiva verificando $\lim\limits_{n\to\infty} a_n = \infty$ e para quase todo o $\omega \in \Omega$ existe um número não-negativo $C = C(\omega)$ tal que,

$$\limsup_{n\to\infty} \Big[a_n \rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)(\omega)\Big] \leqslant C(\omega) \tag{6.2.2}$$

então da anterior fórmula (6.2.1) podemos escrever,

$$\begin{aligned} \|\boldsymbol{\Theta}_n - \boldsymbol{g}_\infty\| &\leqslant \|\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\|\ \|\mathbf{e}_n\| + \|\boldsymbol{g}_n(\boldsymbol{\mu}_n) - \boldsymbol{g}_\infty\| = \\ &= \sqrt{\rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)}\ \|\mathbf{e}_n\| + \|\boldsymbol{g}_n(\boldsymbol{\mu}_n) - \boldsymbol{g}_\infty\| \leqslant \sqrt{C(\omega)}\,\frac{\|\mathbf{e}_n\|}{\sqrt{a_n}} + \|\boldsymbol{g}_n(\boldsymbol{\mu}_n) - \boldsymbol{g}_\infty\| \quad \text{q.c.} \end{aligned} \tag{6.2.3}$$

Em consequência, para vermos que a sucessão de estimadores $\{\boldsymbol{\Theta}_n,\, n \geqslant 1\}$ é fortemente consistente será suficiente mostrar que,

$$\lim_{n\to\infty} \frac{\|\mathbf{e}_n\|^2}{a_n} = 0 \quad \text{quase certamente.}$$

Repare-se que isto não é mais do que estabelecer uma lei forte de grandes números para a sucessão $\mathrm{e}_1, \mathrm{e}_2, \ldots$ de v.a's i.i.d. uma vez que,

$$\frac{\|\mathbf{e}_n\|^2}{a_n} = \frac{\mathrm{e}_1^2 + \ldots + \mathrm{e}_n^2}{a_n}.$$

Para sucessões $\{a_n\}$ verificando $a_n = n^p$ para algum $p \in \left]\frac{1}{2}, \infty\right[$ a lei forte de Marcinkiewicz-Zygmund (conf. Teorema 1.1.17) dá-nos um primeiro resultado para erros não-degenerados, ou seja, erros tais que $\mathbb{E}(\mathrm{e}^2) \neq 0$. Com efeito, o referido resultado diz-nos que dada uma sucessão de variáveis aleatórias $\{\mathrm{X}_n,\, n \geqslant 1\}$ i.i.d. e $q \in]0, 2[$ então,

$$\frac{\mathrm{X}_1 + \ldots + \mathrm{X}_n}{n^{1/q}}$$

converge quase certamente se e só se $\mathbb{E}\,|\mathrm{X}_1|^q < \infty$. Neste caso, o limite é igual a $\mathbb{E}(\mathrm{X}_1)$ se $1 \leqslant q < 2$ ou é nulo se $0 < q < 1$. Consequentemente, para erros não-degenerados $\|\mathbf{e}_n\|^2 / n^p$ convergirá para zero se e só se $p > 1$ e $\mathbb{E}(\mathrm{e}^{2/p}) < \infty$. Deste modo, usando o procedimento de linearização (6.2.1) para a sucessão de estimadores $\{\boldsymbol{\Theta}_n,\, n \geqslant 1\}$ obtemos um primeiro resultado.

Teorema 6.2.1 *Admitamos que no modelo* (6.0.2)*,*

(a) $\boldsymbol{g}_n(\boldsymbol{\mu}_n) \longrightarrow \boldsymbol{g}_\infty$ *quando* $n \to \infty$.

(b) *Existe uma sucessão numérica* $\{a_n\}$ *positiva verificando* $\lim_{n\to\infty} a_n = \infty$, $\dfrac{1}{a_n} = O\left(n^{-p}\right)$, $n \to \infty$ *com* $p > 1$ *e para quase todo o* $\omega \in \Omega$ *existe um número não-negativo* $\mathrm{C} = \mathrm{C}(\omega)$ *tal que,*

$$\limsup_{n\to\infty} \left[a_n \rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)(\omega)\right] \leqslant \mathrm{C}(\omega).$$

(c) *os erros* $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d. tais que* $\mathbb{E}(\mathrm{e}^{2/p}) < \infty$.

Então $\boldsymbol{\Theta}_n \xrightarrow{\text{q.c.}} \boldsymbol{g}_\infty$.

Demonstração. Da fórmula (6.2.3) tem-se,

$$\|\boldsymbol{\Theta}_n - \boldsymbol{g}_\infty\| \leqslant \sqrt{\mathrm{C}(\omega)}\,\frac{\|\mathbf{e}_n\|}{\sqrt{a_n}} + \|\boldsymbol{g}_n(\boldsymbol{\mu}_n) - \boldsymbol{g}_\infty\| \quad \text{q.c.}$$

pelo que a tese é consequência da lei forte de Marcinkiewicz-Zygmund (conf. Teorema 1.1.17) que assegura,

$$\lim_{n\to\infty} \frac{\|\mathbf{e}_n\|^2}{a_n} \leqslant c_3 \lim_{n\to\infty} \frac{\|\mathbf{e}_n\|^2}{n^p} = 0$$

para alguma constante $c_3 > 0$. ∎

Os resultados seguintes exploram as observações efectuadas acima em conjunção com o lema de Kronecker (conf. Proposição 1.3.2) e o teorema das três séries de Kolmogorov (conf. Teorema 1.1.19). Os resultados podem aplicar-se a erros cuja distribuição tem DEI.

Teorema 6.2.2 *Suponhamos que no modelo* (6.0.2),

(a) $\boldsymbol{g}_n(\boldsymbol{\mu}_n) \longrightarrow \boldsymbol{g}_\infty$ *quando* $n \to \infty$.

(b) *Existe uma sucessão numérica* $\{a_n\}$ *positiva verificando* $\lim\limits_{n\to\infty} a_n = \infty$ *e para quase todo o* $\omega \in \Omega$ *existe um número não-negativo* $\mathrm{C} = \mathrm{C}(\omega)$ *tal que,*

$$\limsup_{n\to\infty} \Big[a_n \rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)(\omega)\Big] \leqslant \mathrm{C}(\omega).$$

(c) *Os erros* $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \dots$ *são i.i.d.*

(d) $\sum\limits_{n=1}^{\infty} \mathbb{P}\{\mathrm{e}^2 > a_n\} < \infty$, $\sum\limits_{n=1}^{\infty} \dfrac{1}{a_n} \mathbb{E}\left(\mathrm{e}^2 I_{\{\mathrm{e}^2 \leqslant a_n\}}\right) < \infty$ *e* $\sum\limits_{n=1}^{\infty} \dfrac{1}{a_n^2} \mathbb{E}\left(\mathrm{e}^4 I_{\{\mathrm{e}^2 \leqslant a_n\}}\right) < \infty$.

Então $\boldsymbol{\Theta}_n \xrightarrow{\text{q.c.}} \boldsymbol{g}_\infty$.

Demonstração. No contexto em que desenvolvemos este trabalho, se definirmos

$$\mathrm{Z}_n = \frac{\mathrm{e}_n^2}{a_n} \quad \text{e} \quad \mathrm{Z}'_n = \frac{\mathrm{e}_n^2}{a_n} I_{\left\{\frac{\mathrm{e}_n^2}{a_n} \leqslant 1\right\}}$$

então a hipótese (d) pode ser reescrita como,

$$\sum_{n=1}^{\infty} \mathbb{P}\{\mathrm{Z}_n > 1\} < \infty, \quad \sum_{n=1}^{\infty} \mathbb{E}(\mathrm{Z}'_n) < \infty \quad \text{e} \quad \sum_{n=1}^{\infty} \mathbb{V}\left(\mathrm{Z}'_n\right) < \infty$$

o que pelo teorema das três séries de Kolmogorov (conf. Teorema 1.1.19) é equivalente à convergência quase certa de

$$\sum_{n=1}^{\infty} \mathrm{Z}_n = \sum_{n=1}^{\infty} \frac{\mathrm{e}_n^2}{a_n}.$$

O lema de Kronecker garante então que $\lim\limits_{n\to\infty} \dfrac{\|\mathbf{e}_n\|^2}{a_n} = 0$ quase certamente, pelo que a tese anunciada é agora consequência das restantes hipóteses assumidas e da majoração dada por (6.2.3). ∎

O resultado seguinte dá condições mais simples para a convergência forte da sucessão de estimadores $\{\boldsymbol{\Theta}_n,\, n \geqslant 1\}$ quando os erros têm DEI.

Corolário 6.2.1 *Admitamos que no modelo* (6.0.2),

(a) $\boldsymbol{g}_n(\boldsymbol{\mu}_n) \longrightarrow \boldsymbol{g}_\infty$ *quando* $n \to \infty$.

(b) *Existe uma sucessão numérica* $\{a_n\}$ *positiva verificando* $\lim\limits_{n\to\infty} a_n = \infty$, $\sum\limits_{n=1}^{\infty} \dfrac{1}{a_n} < \infty$ *e para quase todo o* $\omega \in \Omega$ *existe um número não-negativo* $\mathrm{C} = \mathrm{C}(\omega)$ *tal que,*

$$\limsup_{n\to\infty} \Big[a_n \rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)(\omega)\Big] \leqslant \mathrm{C}(\omega).$$

(c) *Os erros* $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d. tais que* $\mathrm{e} \sim \mathrm{DEI}(\alpha, \beta, \tau)$.

Então $\boldsymbol{\Theta}_n \xrightarrow{\text{q.c.}} \boldsymbol{g}_\infty$.

Demonstração. Demonstremos a hipótese (d) do Teorema 6.2.2. A desigualdade de Markov (conf. Observação 1.1.5) garante que,

$$\mathbb{P}\{\mathrm{e}^2 > a_n\} \leqslant \frac{\mathbb{E}(\mathrm{e}^2)}{a_n}$$

e a condição $\sum_{n=1}^{\infty} \frac{1}{a_n} < \infty$ estabelece $\sum_{n=1}^{\infty} \mathbb{P}\{\mathrm{e}^2 > a_n\} < \infty$. Por outro lado,

$$\mathbb{E}\big(\mathrm{e}^2 I_{\{\mathrm{e}^2 \leqslant a_n\}}\big) \leqslant \mathbb{E}\big(\mathrm{e}^2\big) < \infty$$

e

$$\mathbb{E}\big(\mathrm{e}^4 I_{\{\mathrm{e}^2 \leqslant a_n\}}\big) \leqslant \mathbb{E}\big(\mathrm{e}^4\big) < \infty$$

o que implica,

$$\sum_{n=1}^{\infty} \frac{1}{a_n} \mathbb{E}\big(\mathrm{e}^2 I_{\{\mathrm{e}^2 \leqslant a_n\}}\big) \leqslant \mathbb{E}\big(\mathrm{e}^2\big) \sum_{n=1}^{\infty} \frac{1}{a_n} < \infty$$

e

$$\sum_{n=1}^{\infty} \frac{1}{a_n^2} \mathbb{E}\big(\mathrm{e}^4 I_{\{\mathrm{e}^2 \leqslant a_n\}}\big) \leqslant \mathbb{E}\big(\mathrm{e}^4\big) \sum_{n=1}^{\infty} \frac{1}{a_n^2} < \infty. \qquad \blacksquare$$

O resultado seguinte aplica-se sempre que os erros tem um momento de segunda ordem finito.

Teorema 6.2.3 *Admitamos que no modelo* (6.0.2),

(a) $\boldsymbol{g}_n(\boldsymbol{\mu}_n) \longrightarrow \boldsymbol{g}_\infty$ *quando* $n \to \infty$.

(b) *Existe uma sucessão numérica* $\{a_n\}$ *positiva verificando* $\lim_{n\to\infty} a_n = \infty$, $\sum_{n=1}^{\infty} \frac{1}{a_n} < \infty$ *e para quase todo o* $\omega \in \Omega$ *existe um número não-negativo* $\mathrm{C} = \mathrm{C}(\omega)$ *tal que,*

$$\limsup_{n\to\infty} \Big[a_n \rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)(\omega)\Big] \leqslant \mathrm{C}(\omega).$$

(c) *Os erros* $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d. tais que* $\mathbb{E}(\mathrm{e}^2) < \infty$.

Então $\boldsymbol{\Theta}_n \xrightarrow{\text{q.c.}} \boldsymbol{g}_\infty$.

Demonstração. Visto que,

$$\sum_{n=1}^{\infty} \mathbb{E}\left(\frac{\mathrm{e}_n^2}{a_n}\right) \leqslant \mathbb{E}(\mathrm{e}^2) \sum_{n=1}^{\infty} \frac{1}{a_n} < \infty$$

obtém-se $\sum_{n=1}^{\infty} \frac{\mathrm{e}_n^2}{a_n} < \infty$ quase certamente[1] e o lema de Kronecker (conf. Proposição 1.3.2) assegura que $\|\mathbf{e}_n\|^2 / a_n \xrightarrow{\text{q.c.}} 0$. A conclusão segue de (6.2.3). $\blacksquare$

[1]Qualquer série $\sum_{n=1}^{\infty} \mathrm{X}_n$ de v.a.'s converge absolutamente q.c. se $\sum_{n=1}^{\infty} \mathbb{E}\,|\mathrm{X}_n|^r < \infty$ para algum $r \in]0,1]$ (ver [55]).

Até aqui, temo-nos baseado no lema de Kronecker para demonstrar que

$$\lim_{n\to\infty} \frac{\|\mathbf{e}_n\|^2}{a_n} = 0 \quad \text{q.c.}$$

à custa da aplicação da lei forte de Marcinkiewicz-Zygmund ou do teorema das três séries de Kolmogorov. No que se vai seguir, vamos fazer uso de outro resultado para atingir o mesmo fim, mais concretamente da lei forte de Feller (conf. Teorema 1.1.20), ultrapassando assim a necessidade do lema de Kronecker.

Teorema 6.2.4 *Suponhamos que no modelo* (6.0.2)*,*

(a) $\boldsymbol{g}_n(\boldsymbol{\mu}_n) \longrightarrow \boldsymbol{g}_\infty$ *quando* $n \to \infty$.

(b) *Existe uma sucessão numérica* $\{a_n\}$ *positiva verificando* $\lim\limits_{n\to\infty} \dfrac{a_n}{n} = \infty$, $\dfrac{a_n}{n}$ *crescente,* $\sum\limits_{n=1}^{\infty} \mathbb{P}\{\mathrm{e}^2 > a_n\} < \infty$ *e para quase todo o* $\omega \in \Omega$ *existe um número não-negativo* $\mathrm{C} = \mathrm{C}(\omega)$ *tal que,*

$$\limsup_{n\to\infty} \Big[a_n \rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)(\omega)\Big] \leqslant \mathrm{C}(\omega).$$

(c) *Os erros* $\mathrm{e}, \mathrm{e}_1, \mathrm{e}_2, \ldots$ *são i.i.d.*

Então $\boldsymbol{\Theta}_n \xrightarrow{\text{q.c.}} \boldsymbol{g}_\infty$.

Demonstração. Pela lei forte de Feller (conf. Teorema 1.1.20) as condições do enunciado garantem que,

$$\lim_{n\to\infty} \frac{\|\mathbf{e}_n\|^2}{a_n} = 0 \quad \text{q.c.}$$

e a tese fica estabelecida através da estimativa (6.2.3). ∎

6.3 Na simetria radial de erros independentes

Em termos gerais, postular que o vector dos erros $\mathbf{e}_n$ tem simetria radial (isto é, que a função de densidade conjunta de $\mathbf{e}_n$ depende apenas da distância à origem através de alguma função não-negativa g) é uma hipótese natural com significado claro, nomeadamente, que não há qualquer direcção privilegiada para os erros em $\mathbb{R}^n$. Geometricamente, as transformações ortogonais correspondem a rotações quando aplicadas a um vector, que mantêm a norma deste, originando a sua permanência sobre a mesma esfera de $\mathbb{R}^n$ centrada na origem. Uma vez que esta esfera é, na realidade, a superfície de nível da densidade do vector dos erros quando este possui simetria radial, conclui-se que esta superfície de nível é invariante para as transformações ortogonais. Por outro lado, a suposição de que o vector dos erros tem simetria radial traduz matematicamente a hipótese da não existência de erros sistemáticos[2] numa determinada experiência. Contudo, assumir que $\mathbf{e}_n$ tem simetria radial, dada a amostra $\mathbf{Y}_n = \boldsymbol{\mu}_n + \mathbf{e}_n$, não constitui uma verdadeira extensão do caso em que se admite uma distribuição normal para os erros independentes. Efectivamente, vimos no capítulo 4 que quando os erros são independentes e $\mathbf{e}_n = (\mathrm{e}_1, \ldots, \mathrm{e}_n)$ tem simetria radial com função g diferenciável ou densidades contínuas para cada e_i então necessariamente os erros $\mathrm{e}_1, \ldots, \mathrm{e}_n$ tem distribuição normal com média nula e igual variância (conf. Teorema 4.2.4 e Teorema 4.2.5). Mais geralmente, é sabido que se $\mathbb{p}$ é uma medida de probabilidade em $\mathbb{R}^n$ invariante relativamente às transformações ortogonais e é, em simultâneo, o produto de duas medidas concentradas em subespaços ortogonais não triviais de $\mathbb{R}^n$ então $\mathbb{p}$ é a medida de Dirac ou $\mathbb{p}$ é a medida Gaussiana com densidade $\boldsymbol{x} \mapsto \frac{1}{(4\pi\sigma)^{n/2}} \exp\left(-\frac{\|\boldsymbol{x}\|^2}{4\sigma}\right)$ para algum $\sigma > 0$ (ver [60]). Uma vez que o vector aleatório $\mathbf{e}_n$, expresso na base canónica de $\mathbb{R}^n$, tem como componentes v.a.'s independentes, existe uma decomposição óbvia da distribuição conjunta de $\mathbf{e}_n$ no produto de distribuições concentradas em subespaços ortogonais não triviais donde um vector aleatório com componentes independentes e simetria radial tem necessariamente distribuição normal multivariada ou degenerada. Deste modo, ao assumirmos que $\mathbf{e}_n$ tem simetria radial e também componentes independentes, estaremos a admitir que os erros são gaussianos já que, na prática, o caso degenerado não possui grande interesse.

Definindo $\mathbb{D}_n$ como o espaço imagem aleatório da aplicação aleatória associada à matriz $\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)^T$ obtém-se, utilizando a Proposição 1.2.4 com $\boldsymbol{A} = \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)$ e $\boldsymbol{x} = \mathbf{e}_n$, que

$$\|\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\mathbf{e}_n\|^2 \leqslant \rho\Big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)^T\Big)\, \|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2$$

donde podemos reescrever a fórmula (6.2.3) do seguinte modo,

$$\|\boldsymbol{\Theta}_n - \boldsymbol{g}_\infty\| \leqslant \sqrt{\mathrm{C}(\omega)}\, \frac{\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|}{\sqrt{a_n}} + \|\boldsymbol{g}_n(\boldsymbol{\mu}_n) - \boldsymbol{g}_\infty\| \tag{6.3.1}$$

para alguma sucessão de reais $\{a_n\}$ positiva. Refira-se que

$$s = \dim(\mathbb{D}_n) = \mathrm{car}\left(\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)^T\right) \leqslant \min\{\kappa, n\} = \kappa \qquad (n \geqslant \kappa)$$

(conf. Teorema 1.2.4) donde se assumirmos que os erros são independentes com simetria radial então o teorema de Cochran (ver Teorema 1.2.10) garante que a distribuição de $\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2$ é independente de n o que permite um melhoramento do Teorema 6.2.3.

Para cada $n \in \mathbb{N}$, considere-se o vector aleatório $\mathbf{e}_n$ e denotemos a respectiva projecção ortogonal no subespaço aleatório $\mathbb{D}_n$ (cuja dimensão é $s \leqslant \kappa$) por $\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n$.

Proposição 6.3.1 *Seja* $\{\mathrm{e}_n,\, n \geqslant 1\}$ *uma sucessão de variáveis aleatórias i.i.d. com distribuição* $N(0, \sigma^2)$. *Se* $\{a_n\}$ *é uma sucessão de reais positiva tal que* $\lim\limits_{n\to\infty} a_n = \infty$ *e* $\sum_{n=1}^{\infty} e^{-\delta a_n} < \infty$ *para*

[2] Na categoria dos erros experimentais encontram-se os **erros sistemáticos** que se caracterizam por afectar sempre no mesmo sentido e com magnitudes que pouco variam o resultado de uma medição. Os erros sistemáticos classificam-se em quatro tipos: instrumentais, de paralaxe, ambientais ou teóricos. Estes erros são dos mais graves pois são frequentemente difíceis de detectar.

qualquer $\delta > 0$ então

$$\frac{\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2}{a_n} \xrightarrow{\text{q.c.}} 0.$$

Demonstração. Do teorema de Cochran e das hipóteses assumidas tem-se $\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2 \sim \sigma^2\chi^2(s)$, $\forall n \in \mathbb{N}$ com $s \leqslant \kappa$ e consequentemente $\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2 \sim \text{DEI}\left(1, \frac{1}{2}, \frac{s}{2} - 1\right)$ pelo que a conclusão decorre da Proposição 6.1.2. ∎

Vamos agora anunciar e demonstrar um resultado geral para erros aleatórios gaussianos independentes.

Teorema 6.3.1 *Suponhamos que no modelo* (6.0.2)*,*

(a) $\boldsymbol{g}_n(\boldsymbol{\mu}_n) \longrightarrow \boldsymbol{g}_\infty$ *quando* $n \to \infty$.

(b) *Existe uma sucessão numérica* $\{a_n\}$ *positiva verificando* $\lim\limits_{n\to\infty} a_n = \infty$, $\sum_{n=1}^{\infty} e^{-\delta a_n} < \infty$ *para qualquer* $\delta > 0$ *e para quase todo o* $\omega \in \Omega$ *existe um número não-negativo* $\mathrm{C} = \mathrm{C}(\omega)$ *tal que,*

$$\limsup_{n\to\infty} \left[a_n \rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T \mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)(\omega)\right] \leqslant \mathrm{C}(\omega).$$

(c) $\{\mathrm{e}_n,\, n \geqslant 1\}$ *é uma sucessão de variáveis aleatórias i.i.d. com distribuição* $N(0, \sigma^2)$.

Então $\boldsymbol{\Theta}_n \xrightarrow{\text{q.c.}} \boldsymbol{g}_\infty$.

Demonstração. De acordo com a Proposição 6.3.1 tem-se,

$$\frac{\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2}{a_n} \xrightarrow{\text{q.c.}} 0$$

e a fórmula (6.3.1) estabelece a tese uma vez que a hipótese (a) assegura a convergência de $\boldsymbol{g}_n(\boldsymbol{\mu}_n)$ para $\boldsymbol{g}_\infty$. ∎

Quando a aplicação diferencial (aleatória) $\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)$ possui entradas que são independentes dos erros ocorre um curioso fenómeno análogo ao verificado no Capítulo 3 (ver Teorema 3.1.1) e que sugestivamente podemos apelidar de *colapso dos erros*.

Proposição 6.3.2 *Se* $\{\mathrm{e}_n,\, n \geqslant 1\}$ *é uma sucessão de variáveis aleatórias i.i.d. com distribuição* $N(0, \sigma^2)$ *e cada sucessão de v.a's* $\left\{\left[\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\right]_{ij},\ j \geqslant 1\right\}$ $(i = 1, \ldots, \kappa)$ *é independente da sucessão* $\{\mathrm{e}_n,\, n \geqslant 1\}$ *então para qualquer sucessão de reais positiva* $\{a_n\}$ *satisfazendo* $\lim\limits_{n\to\infty} a_n = \infty$ *tem-se,*

$$\frac{\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2}{a_n} \xrightarrow{\text{q.c.}} 0.$$

Demonstração. Aplicando a Proposição 1.2.1 a $\boldsymbol{A} = \big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)^T$ e $\boldsymbol{x} = \mathbf{e}_n$ existirá uma base (aleatória) ortonormada

$$\begin{aligned}\mathbf{W}_1(n,\omega) &= \big(\mathrm{W}_{11}(n,\omega),\ldots,\mathrm{W}_{n1}(n,\omega)\big)\\ &\vdots\\ \mathbf{W}_s(n,\omega) &= \big(\mathrm{W}_{1\kappa}(n,\omega),\ldots,\mathrm{W}_{ns}(n,\omega)\big)\end{aligned}$$

tal que a v.a. $\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2$ é expressa por,

$$\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2 = \Big(\mathrm{S}_{n1}(n)\Big)^2 + \ldots + \Big(\mathrm{S}_{ns}(n)\Big)^2 \tag{6.3.2}$$

onde a sucessão $\big\{\mathrm{S}_{ni}(n),\, n \geqslant 1\big\}$ está definida pelo arranjo triangular de v.a.'s,

$$\begin{aligned}\mathrm{S}_{1i}(1) &= \mathrm{W}_{1i}(1,\omega)\mathrm{e}_1\\ \mathrm{S}_{2i}(2) &= \mathrm{W}_{1i}(2,\omega)\mathrm{e}_1 + \mathrm{W}_{2i}(2,\omega)\mathrm{e}_2\\ &\vdots\\ \mathrm{S}_{ni}(n) &= \mathrm{W}_{1i}(n,\omega)\mathrm{e}_1 + \mathrm{W}_{2i}(n,\omega)\mathrm{e}_2 + \ldots + \mathrm{W}_{ni}(n,\omega)\mathrm{e}_n\end{aligned}$$

para todo o $i = 1,\ldots,s$. Visto que,

$$\begin{aligned}&\mathbb{E}\big[\mathrm{W}_{1i}(1,\omega)\mathrm{e}_1\big] = \mathbb{E}\big[\mathrm{W}_{1i}(1,\omega)\big]\cdot\mathbb{E}(\mathrm{e}_1) = 0 \quad \text{q.c.}\\ &\mathbb{E}\big[\mathrm{W}_{2i}(2,\omega)\mathrm{e}_2 \mid \mathrm{W}_{1i}(2,\omega)\mathrm{e}_1\big] = \mathrm{W}_{2i}(2,\omega)\mathbb{E}\big[\mathrm{e}_2 \mid \mathrm{W}_{1i}(2,\omega)\mathrm{e}_1\big] = 0 \quad \text{q.c.}\\ &\qquad\vdots\\ &\mathbb{E}\big[\mathrm{W}_{ni}(n,\omega)\mathrm{e}_n \mid \mathrm{W}_{1i}(n,\omega)\mathrm{e}_1,\ldots,\mathrm{W}_{(n-1)i}(n,\omega)\mathrm{e}_{n-1}\big] =\\ &\qquad = \mathrm{W}_{ni}(n,\omega)\mathbb{E}\big[\mathrm{e}_n \mid \mathrm{W}_{1i}(n,\omega)\mathrm{e}_1,\ldots,\mathrm{W}_{(n-1)i}(n,\omega)\mathrm{e}_{n-1}\big] = 0 \quad \text{q.c.}\end{aligned}$$

podemos aplicar a igualdade extendida de Bienaymé e a desigualdade extendida de Kolmogorov (ver Teorema 1.1.13 e Teorema 1.1.14) a cada termo da sucessão aleatória $\big\{\mathrm{S}_{ni}(n),\, n \geqslant 1\big\}$ o que produz,[3]

$$\begin{aligned}\mathbb{P}\left\{\max_{1\leqslant j\leqslant 1}|\mathrm{S}_{ji}(1)| \geqslant \varepsilon\right\} &\leqslant \frac{1}{\varepsilon^2}\cdot\mathbb{E}\left[\Big(\mathrm{S}_{1i}(1)\Big)^2\right] \leqslant \frac{\mathbb{E}\Big(\|\boldsymbol{P}_{\mathbb{D}_1}\mathbf{e}_1\|^2\Big)}{\varepsilon^2} = \frac{2s(s+2)\sigma^2}{\varepsilon^2}\\ \mathbb{P}\left\{\max_{1\leqslant j\leqslant 2}|\mathrm{S}_{ji}(2)| \geqslant \varepsilon\right\} &\leqslant \frac{1}{\varepsilon^2}\cdot\mathbb{E}\left[\Big(\mathrm{S}_{2i}(2)\Big)^2\right] \leqslant \frac{\mathbb{E}\Big(\|\boldsymbol{P}_{\mathbb{D}_2}\mathbf{e}_2\|^2\Big)}{\varepsilon^2} = \frac{2s(s+2)\sigma^2}{\varepsilon^2}\\ &\vdots\\ \mathbb{P}\left\{\max_{1\leqslant j\leqslant n}|\mathrm{S}_{ji}(n)| \geqslant \varepsilon\right\} &\leqslant \frac{1}{\varepsilon^2}\cdot\mathbb{E}\left[\Big(\mathrm{S}_{ni}(n)\Big)^2\right] \leqslant \frac{\mathbb{E}\Big(\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2\Big)}{\varepsilon^2} = \frac{2s(s+2)\sigma^2}{\varepsilon^2}\end{aligned}$$

para qualquer $i = 1,\ldots,s$ pois o teorema de Cochran (ver Teorema 1.2.10) garante que $\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2 \sim 2\sigma^2\chi^2(s)$, $\forall n \in \mathbb{N}$. Então,

$$\mathbb{P}\left\{\max_{1\leqslant j\leqslant n}|\mathrm{S}_{ji}(n)| \geqslant \varepsilon\right\} \leqslant \frac{2s(s+2)\sigma^2}{\varepsilon^2}, \quad \forall n \in \mathbb{N}$$

[3] Note-se que $\mathbb{E}\big[\mathrm{S}_{ni}(n)\big] = 0$, $\forall n \geqslant 1$.

e escolhendo a subsucessão $\max\limits_{1\leqslant j\leqslant \xi_n} |\mathrm{S}_{ji}(\xi_n)|$ de $\max\limits_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)|$ que nos dá,

$$\lim_{n\to\infty}\left(\max_{1\leqslant j\leqslant \xi_n} |\mathrm{S}_{ji}(\xi_n)|\right) = \limsup_{n\to\infty}\left(\max_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)|\right)$$

vem,

$$\mathbb{P}\left\{\max_{1\leqslant j\leqslant \xi_n} |\mathrm{S}_{ji}(\xi_n)| \geqslant \varepsilon\right\} \leqslant \frac{2s(s+2)\sigma^2}{\varepsilon^2}, \quad \forall n \in \mathbb{N}.$$

Efectuando a passagem ao limite em n surge,

$$\mathbb{P}\left\{\limsup_{n\to\infty}\left(\max_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)|\right) \geqslant \varepsilon\right\} \leqslant \frac{2s(s+2)\sigma^2}{\varepsilon^2}$$

e fazendo $\varepsilon \to \infty$ concluí-se que $\limsup\limits_{n\to\infty}\left(\max\limits_{1\leqslant j\leqslant n} |\mathrm{S}_{ji}(n)|\right)$ existe finito quase certamente para todo o $i = 1, \ldots, \kappa$. Consequentemente,

$$\limsup_{n\to\infty}\ \|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2$$

existe finito quase certamente verificando-se ainda $\limsup\limits_{n\to\infty}\ \|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2 > 0$ q.c. uma vez que $\|\boldsymbol{P}_{\mathbb{D}_n}\mathbf{e}_n\|^2 \sim 2\sigma^2\chi^2(s)$, $\forall n \in \mathbb{N}$. ■

Teorema 6.3.2 *Suponhamos que no modelo* (6.0.2),

(a) $\boldsymbol{g}_n(\boldsymbol{\mu}_n) \longrightarrow \boldsymbol{g}_\infty$ *quando* $n \to \infty$.

(b) *Existe uma sucessão numérica* $\{a_n\}$ *positiva verificando* $\lim\limits_{n\to\infty} a_n = \infty$ *e para quase todo o* $\omega \in \Omega$ *existe um número não-negativo* $\mathrm{C} = \mathrm{C}(\omega)$ *tal que,*

$$\limsup_{n\to\infty}\Big[a_n\rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big)(\omega)\Big] \leqslant \mathrm{C}(\omega).$$

(c) $\{\mathrm{e}_n,\, n \geqslant 1\}$ *é uma sucessão de variáveis aleatórias i.i.d. com distribuição* $N(0, \sigma^2)$.

(d) *Cada sucessão de v.a's* $\big\{\big[\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big]_{ij},\ j \geqslant 1\big\}$ $(i = 1, \ldots, \kappa)$ *é independente da sucessão* $\{\mathrm{e}_n,\, n \geqslant 1\}$

Então $\boldsymbol{\Theta}_n \xrightarrow{\text{q.c.}} \boldsymbol{g}_\infty$.

Demonstração. A tese é uma consequência da estimativa (6.3.1) e da Proposição 6.3.2. ■

Observação 6.3.1 Observemos que, se no modelo (6.0.2) a matriz $\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)$ for não-estocástica então as hipóteses (a), (b) e (c) do Teorema 6.3.2 ainda garantem a convergência $\boldsymbol{\Theta}_n \xrightarrow{\text{q.c.}} \boldsymbol{g}_\infty$.

6.4 Exemplos de aplicação

A finalizar, apresentamos dois exemplos de estimadores não-lineares relativamente aos quais a teoria desenvolvida anteriormente se aplica. Mais concretamente, mostramos que se considerarmos pesos adequados para a sucessão das observações, o raio espectral da aplicação diferencial quando multiplicada pela sua transposta pode ter uma velocidade de convergência para zero suficiente para que o estimador seja convergente quase certamente, pelo menos, no caso dos erros de quadrado integrável.

Exemplo 6.4.1
Seja $g_n(x_1,\ldots,x_n) = \cos\left(\dfrac{x_1+\ldots+x_n}{2^n}\right)$ e suponha-se $\mu_n = 2^{n-1}\tau_n$, com $\{\tau_n\}$ uma sucessão convergente tal que $\lim_{n\to\infty}\tau_n = \tau_\infty \in \mathbb{R}$. Pelo lema de Cesàro (conf. Proposição 1.3.1) obtém-se,

$$\lim_{n\to\infty} \frac{\tau_1+\ldots+2^{n-1}\tau_n}{2^n} = \tau_\infty$$

o que implica $\lim\limits_{n\to\infty} g_n(\mu_1,\ldots,\mu_n) = \cos\big(\tau_\infty\big)$. Por outro lado, para algum vector aleatório $\boldsymbol{\lambda}_n = (\lambda_1,\ldots,\lambda_n) \in \mathbb{L}$,

$$\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T = \left[\begin{array}{ccc} -\dfrac{1}{2^n}\sin\left(\dfrac{\lambda_1+\ldots+\lambda_n}{2^n}\right) & \ldots & -\dfrac{1}{2^n}\sin\left(\dfrac{\lambda_1+\ldots+\lambda_n}{2^n}\right)\end{array}\right].$$

Aplicando o Teorema 6.2.3 com $a_n = \dfrac{4^n}{n}$ podemos concluir que $\Theta_n \xrightarrow{\text{q.c.}} \cos\big(\tau_\infty\big)$ sempre que $\mathbb{E}\big(\mathrm{e}^2\big) < \infty$ já que,

$$\rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big) = \frac{1}{4^n}\sin^2\left(\frac{\lambda_1+\ldots+\lambda_n}{2^n}\right) + \ldots + \frac{1}{4^n}\sin^2\left(\frac{\lambda_1+\ldots+\lambda_n}{2^n}\right) \leqslant \frac{n}{4^n}.$$

Exemplo 6.4.2
Consideremos $g_n(x_1,\ldots,x_n) = \log\left(1+\left(\dfrac{x_1+\ldots+x_n}{n^2}\right)^2\right)$ e admita-se que $\mu_n = (2n-1)v_n$, sendo $\{v_n\}$ uma sucessão convergente verificando $\lim\limits_{n\to\infty} v_n = v_\infty \in \mathbb{R}$. Pelo lema de Cesàro,

$$\lim_{n\to\infty} \frac{v_1+\ldots+(2n-1)v_n}{n^2} = v_\infty$$

o que implica que $\lim\limits_{n\to\infty} g_n(\mu_1,\ldots,\mu_n) = \log\big(1+v_\infty^2\big)$. Por outro lado, para algum vector aleatório $\boldsymbol{\lambda}_n = (\lambda_1,\ldots,\lambda_n) \in \mathbb{L}$ tem-se,

$$\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T = \left[\begin{array}{ccc} \dfrac{\frac{2}{n^2}\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)}{1+\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)^2} & \ldots & \dfrac{\frac{2}{n^2}\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)}{1+\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)^2}\end{array}\right]$$

Visto que a função real $u\colon [0,\infty[\longrightarrow \mathbb{R}$ definida por $u(t) = \dfrac{t}{(1+t)^2}$ atinge o valor máximo em $t_0 = 1$ vem,

$$\begin{aligned}\rho\big(\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)^T\mathbf{D}\boldsymbol{g}_n(\boldsymbol{\lambda}_n)\big) &= \frac{\frac{4}{n^4}\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)^2}{\left[1+\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)^2\right]^2} + \ldots + \frac{\frac{4}{n^4}\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)^2}{\left[1+\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)^2\right]^2} = \\ &= \frac{4}{n^3}\cdot\frac{\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)^2}{\left[1+\left(\frac{\lambda_1+\ldots+\lambda_n}{n^2}\right)^2\right]^2} \leqslant \frac{4}{n^3}\cdot\frac{1}{4} = \frac{1}{n^3}\end{aligned}$$

Aplicando o Teorema 6.2.3 com $a_n = n^3$ pode concluir-se a convergência forte do estimador,

$$\Theta_n \xrightarrow{\text{q.c.}} \log\left(1 + v_\infty^2\right)$$

sempre que $\mathbb{E}\left(\mathrm{e}^2\right) < \infty$.

Apêndice

DISTRIBUIÇÃO ARCSIN

Uma variável aleatória X tem distribuição arcsin se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{1}{\pi\sqrt{x(1-x)}} & \text{se} \quad x \in]0,1[\\ \\ 0 & \text{se} \quad x \notin]0,1[\end{cases}$$

Propriedades:

1. $\mathbb{E}\left(\mathrm{X}^k\right) = \frac{\Gamma\left(k+\frac{1}{2}\right)}{\sqrt{\pi}(k+1)}$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = 1/8$.

3. $\mathbb{E}\left(e^{it\mathrm{X}}\right) = \mathcal{B}_0\left(\frac{t}{2}\right)\cos\left(\frac{t}{2}\right) + i\mathcal{B}_0\left(\frac{t}{2}\right)\sin\left(\frac{t}{2}\right)$, $t \in \mathbb{R}$ onde $\mathcal{B}_n(x)$ é a função de Bessel de primeira espécie.[4]

DISTRIBUIÇÃO BETA

Uma variável aleatória X tem distribuição beta de parâmetros (α, β) com $\alpha, \beta > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{se} \quad x \in [0,1] \\ \\ 0 & \text{se} \quad x \notin [0,1] \end{cases}$$

Propriedades:

1. $\mathbb{E}\left(\mathrm{X}^k\right) = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

3. O valor $x_0 = \frac{\alpha-1}{\alpha+\beta-2}$ é a única moda quando $\alpha > 1$ e $\beta > 1$.

4. $\mathbb{E}\left(e^{it\mathrm{X}}\right) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \frac{\Gamma(\alpha+n)}{\Gamma(\alpha+\beta+n)}$, $t \in \mathbb{R}$.

---

[4] $\mathcal{B}_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{m=0}^{\infty} \frac{(-1)^m \left(\frac{z}{2}\right)^{2m}}{\Gamma(m+\nu+1)\, m!}$, $z \in \mathbb{C}$.

DISTRIBUIÇÃO BENKTANDER DO TIPO I

Uma variável aleatória X tem distribuição Benktander do tipo I de parâmetros (α, β) com $\alpha, \beta > 0$ se tiver função de distribuição,

$$F(x) = \begin{cases} 1 - \left(1 + \dfrac{2\beta \log x}{\alpha}\right) e^{-\beta \log^2(x) - (\alpha+1)\log x} & \text{se} \quad x > 1 \\ \\ 0 & \text{se} \quad x \leqslant 1 \end{cases}.$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = 1 + \frac{k}{\alpha} + \frac{k(k-1)\sqrt{\pi}}{2\alpha\sqrt{\beta}}\left[1 - \operatorname{erf}\left(\frac{1-k+\alpha}{2\sqrt{\beta}}\right)\right] e^{\frac{(1+\alpha)^2 + k^2 - 2k(\alpha+1)}{4\beta}}$, $k \in \mathbb{N}$ onde $\operatorname{erf}(x)$ é a função erro.[5]
2. $\mathbb{V}(\mathrm{X}) = -\frac{1}{\alpha^2} + \frac{\sqrt{\pi}}{\alpha\sqrt{\beta}}\left[1 - \operatorname{erf}\left(\frac{\alpha-1}{2\sqrt{\beta}}\right)\right] e^{\frac{(\alpha-1)^2}{4\beta}}$.
3. O valor $x_0 = e^{-\frac{\alpha+1-\sqrt{1+6\beta}}{2\beta}}$ é moda se $\beta \geqslant \frac{\alpha(\alpha^2+3\alpha+2)}{6(\alpha+1)}$; o valor $x_0 = 1$ é moda se $\beta < \frac{\alpha(\alpha^2+3\alpha+2)}{6(\alpha+1)}$.

DISTRIBUIÇÃO BENKTANDER DO TIPO II

Uma variável aleatória X tem distribuição Benktander do tipo II de parâmetros (α, β) com $\alpha > 0$ e $0 < \beta < 1$ se tiver função de distribuição,

$$F(x) = \begin{cases} 1 - e^{\frac{\alpha}{\beta}} x^{-(1-\beta)} e^{-\frac{\alpha}{\beta} x^{\beta}} & \text{se} \quad x > 1 \\ \\ 0 & \text{se} \quad x \leqslant 1 \end{cases}.$$

Propriedades:

1. $\mathbb{E}\left(\mathrm{X}^k\right) = \left(\frac{1}{\beta} - 1\right) e^{\frac{\alpha}{\beta}} \left(\frac{\beta}{\alpha}\right)^{\frac{k+\beta-1}{\beta}} \Gamma\left(\frac{k+\beta-1}{\beta}, \frac{\alpha}{\beta}\right) + e^{\frac{\alpha}{\beta}} \left(\frac{\beta}{\alpha}\right)^{\frac{k+\beta-1}{\beta}} \Gamma\left(\frac{k+2\beta-1}{\beta}, \frac{\alpha}{\beta}\right)$ onde $\Gamma(x,y)$ é a função gama incompleta.[6]
2. $\mathbb{V}(\mathrm{X}) = \left(\beta^{\frac{1}{\beta}} - \beta^{\frac{1+2\beta}{\beta}}\right) \alpha^{-\frac{1+\beta}{\beta}} \Gamma\left(\frac{1+\beta}{\beta}, \frac{\alpha}{\beta}\right) e^{\frac{\alpha}{\beta}} + \alpha^{-\frac{1+\beta}{\beta}} \beta^{\frac{1+2\beta}{\beta}} \Gamma\left(\frac{1+2\beta}{\beta}, \frac{\alpha}{\beta}\right) e^{\frac{\alpha}{\beta}} - \frac{(\alpha+1)^2}{\alpha^2}$.
3. O valor $x_0 = 1$ é moda.

DISTRIBUIÇÃO DE BURR

Uma variável aleatória X tem distribuição de Burr de parâmetros (β, σ, δ) com $\beta, \sigma, \delta > 0$ se tiver função de distribuição,

$$f(x) = \begin{cases} \dfrac{\sigma\delta}{\beta}\left(\dfrac{x-\alpha}{\beta}\right)^{-\sigma-1}\left[1 + \left(\dfrac{x-\alpha}{\beta}\right)^{-\sigma}\right]^{-\delta-1} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}.$$

Propriedades:

[5] $\operatorname{erf}(x) = \dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^x e^{-t^2}\, dt,\ x \in \mathbb{R}$.

[6] $\Gamma(x,y) = \displaystyle\int_y^{\infty} t^{x-1} e^{-t}\, dt,\ x, y > 0$.

1. $\mathbb{E}(\mathrm{X}^k) = \frac{\beta^k \Gamma\left(\frac{-k+\sigma}{\sigma}\right)\Gamma\left(\frac{\delta\sigma+k}{\sigma}\right)}{\Gamma(\delta)}$, $\sigma > k$.

2. $\mathbb{V}(\mathrm{X}) = \frac{\beta^2}{\Gamma^2(\delta)} \left\{\Gamma(\delta)\Gamma\left(1-\frac{2}{\sigma}\right)\Gamma\left(\frac{2}{\sigma}+\delta\right) - \left[\Gamma\left(1-\frac{1}{\sigma}\right)\right]^2 \left[\Gamma\left(\frac{1}{\sigma}+\delta\right)\right]^2\right\}$, $\sigma > 2$.

3. O valor $x_0 = \beta\left(\frac{\sigma\delta-1}{\sigma+1}\right)^{\frac{1}{\sigma}}$ é moda se $\sigma\delta > 1$; o valor $x_0 = 0$ é moda se $\sigma\delta \leqslant 1$.

DISTRIBUIÇÃO DE CAUCHY

Uma variável aleatória X tem distribuição de Cauchy de parâmetros (μ, σ) com $\mu \in \mathbb{R}$ e $\sigma > 0$ se tiver função densidade de probabilidade

$$f(x) = \frac{1}{\pi}\frac{\sigma}{\sigma^2 + (x-\mu)^2}, \quad x \in \mathbb{R}.$$

Propriedades:

1. Todos os momentos são infinitos.

2. $\mathrm{X} \sim \mathcal{S}_1(\sigma, 0, \mu)$.

3. O valor $x_0 = \mu$ é moda e mediana.

4. $\mathbb{E}(e^{it\mathrm{X}}) = e^{i\mu t - \sigma|t|}$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO DE CAUCHY GENERALIZADA

Uma variável aleatória X tem distribuição de Cauchy generalizada de parâmetro $\alpha > 0$ se tiver função densidade de probabilidade

$$f(x) = \frac{1+\alpha}{2\pi}\frac{\sin\left(\frac{\pi}{1+\alpha}\right)}{1+|x|^{1+\alpha}}, \quad x \in \mathbb{R}.$$

Propriedades:

1. Todos os momentos são infinitos.

2. O valor $x_0 = 0$ é moda.

DISTRIBUIÇÃO DE CAUCHY MULTIVARIADA

Um vector aleatório $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ tem distribuição de Cauchy multivariada com parâmetro α se tiver função de densidade conjunta

$$f(x_1, \ldots, x_n) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\pi^{\frac{\alpha+1}{2}}\left(1 + x_1^2 + \ldots + x_n^2\right)^{\frac{\alpha+1}{2}}}, \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

DISTRIBUIÇÃO DE ERLANG

Uma variável aleatória X tem distribuição de Erlang de parâmetros (n, λ) com $n \in \mathbb{N}$ e $\lambda > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = \frac{n(n+1)\ldots(n+k-1)}{\lambda^k}$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = \frac{n}{\lambda^2}$.

3. O valor $x_0 = \frac{n-1}{\lambda}$ é moda.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = \left(\frac{\lambda}{\lambda - it}\right)^n$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO EXPONENCIAL

Uma variável aleatória tem distribuição exponencial de parâmetro $\alpha > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = \frac{k!}{\lambda^k}$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = \frac{1}{\lambda^2}$.

3. O valor $x_0 = 0$ é moda.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = \frac{\lambda}{\lambda - it}$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO DE FRÉCHET

Uma variável aleatória X tem distribuição de Fréchet de parâmetros (α, β) com $\alpha, \beta > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{\alpha}{\beta}\left(\dfrac{x}{\beta}\right)^{-\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{-\alpha}} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

Propriedades:

1. $\mathbb{E}\left(\mathrm{X}^k\right) = \beta^k \Gamma\left(\frac{\alpha - k}{\alpha}\right)$, $\alpha > k$.

2. $\mathbb{V}(\mathrm{X}) = \beta^2 \left\{\Gamma\left(\frac{\alpha-2}{\alpha}\right) - \left[\Gamma\left(\frac{\alpha-1}{\alpha}\right)\right]^2\right\}$, $\alpha > 2$.

3. O valor $x_0 = \beta\left(\frac{\alpha}{1+\alpha}\right)^{\frac{1}{\alpha}}$ é a única moda.

DISTRIBUIÇÃO GAMA

Uma variável aleatória X tem distribuição gama de parâmetros (α, λ) com $\alpha, \lambda > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = \frac{\alpha(\alpha+1)\ldots(\alpha+k-1)}{\lambda^k}$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = \frac{\alpha}{\lambda^2}$.

3. O valor $x_0 = \frac{\alpha-1}{\lambda}$ é moda quando $\alpha \geqslant 1$.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = \big(1 - \frac{it}{\lambda}\big)^{-\alpha}$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO DE GUMBEL

Uma variável aleatória X tem distribuição de Gumbel de parâmetros (α, β) com $\alpha \in \mathbb{R}$ e $\beta > 0$ se tiver função densidade de probabilidade

$$f(x) = \frac{1}{\beta} e^{-\frac{x-\alpha}{\beta}} e^{\left[-e^{-\frac{x-\alpha}{\beta}}\right]}, \quad x \in \mathbb{R}.$$

Propriedades:

1. $\mathbb{E}\left(\mathrm{X}^k\right) = \sum_{j=0}^{k} \binom{k}{j} \alpha^{k-j} \beta^j \iota_k$, $k \in \mathbb{N}$ onde ι_k é o integral de Euler-Mascheroni.[7]

2. $\mathbb{V}(\mathrm{X}) = \frac{\pi^2 \beta^2}{6}$.

3. O valor $x_0 = \alpha$ é moda.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = e^{\alpha it} \Gamma(1 - \beta it)$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO DE KAPTEYN

Uma variável aleatória X tem distribuição de Kapteyn de parâmetros $(\mu, \sigma^2, G(x))$ com $\mu \in \mathbb{R}$, $\sigma > 0$ e G uma função monótona diferenciável se tiver função densidade de probabilidade

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{[G(x)-\mu]^2}{2\sigma^2}} \left|G'(x)\right|.$$

DISTRIBUIÇÃO DE KOTZ

Um vector aleatório $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ tem distribuição de Kotz de parâmetros (a, b, n, q) com $a, b > 0$ e $n + 2q > 2$ se tiver função densidade de probabilidade conjunta,

$$f(x_1, \ldots, x_n) = \frac{a\, \Gamma\left(\dfrac{n}{2}\right)}{b^{-\frac{2q+n-2}{2a}}\, \pi^{\frac{n}{2}}\, \Gamma\left(\dfrac{2q+n-2}{2a}\right)} (x_1^2 + \ldots + x_n)^{q-1} e^{-b\,(x_1^2+\ldots+x_n^2)^a}$$

[7] $\iota_k = (-1)^k \int_0^\infty (\log t)^k e^{-t}\, dt$.

DISTRIBUIÇÃO DE LÉVY

Uma variável aleatória X tem distribuição de Lévy de parâmetros (μ, σ) com $\mu \in \mathbb{R}$ $\sigma > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \sqrt{\dfrac{\sigma}{2\pi}} \dfrac{1}{(x-\mu)^{3/2}} e^{-\frac{\sigma}{2(x-\mu)}} & \text{se} \quad x > \mu \\ \\ 0 & \text{se} \quad x \leqslant \mu \end{cases}$$

Propriedades:

1. Todos os momentos são infinitos.

2. $\mathrm{X} \sim \mathcal{S}_{1/2}(\sigma, 1, \mu)$.

3. O valor $x_0 = \mu + \frac{\sigma}{3}$ é moda.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = e^{-\left(it\mu + \sqrt{-2i\sigma t}\right)}$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO LOG-GAMA

Uma variável aleatória X tem distribuição log-gama de parâmetros (α, β) com $\alpha, \beta > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{\alpha^\beta}{\Gamma(\beta)} (\log x)^{\beta-1} x^{-\alpha-1} & \text{se} \quad x > 1 \\ \\ 0 & \text{se} \quad x \leqslant 1 \end{cases}$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = \alpha^\beta(\alpha-k)^{-\beta}$, $\alpha > k$.

2. $\mathbb{V}(\mathrm{X}) = \alpha^\beta(\alpha-2)^{-\beta} - \alpha^{2\beta}(\alpha-1)^{-2\alpha}$.

3. O valor $x_0 = e^{\left(\frac{\beta-1}{\alpha+1}\right)}$ é moda se $\beta \geqslant 1$.

DISTRIBUIÇÃO LOG-NORMAL

Uma variável aleatória X tem distribuição log-normal de parâmetros (μ, σ^2) com $\mu \in \mathbb{R}$ e $\sigma > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = e^{k\mu + \frac{k^2\sigma^4}{2}}$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = e^{\sigma^2+\mu}\big[e^{\sigma^4} - 1\big]$.

3. O valor $x_0 = e^{\mu-\sigma^2}$ é a única moda.

DISTRIBUIÇÃO DE MAXWELL

Uma variável aleatória X tem distribuição de Maxwell de parâmetro $\alpha > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \sqrt{\dfrac{2}{\pi}}\dfrac{x^2 e^{-\frac{x^2}{2\alpha^2}}}{\alpha^3} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = \frac{2^{1+\frac{k}{2}}\alpha^k\Gamma\left(\frac{k+3}{2}\right)}{\sqrt{\pi}}$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = \frac{\alpha^2(3\pi-8)}{\pi}$.

3. O valor $x_0 = \alpha\sqrt{2}$ é moda.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = i\left\{\alpha t\sqrt{\frac{2}{\pi}} + ie^{-\frac{\alpha^2t^2}{2}}(\alpha^2t^2-1)\left[\mathrm{sign}(t)\mathrm{erf}\left(\frac{\alpha|t|i}{\sqrt{2}}\right)+1\right]\right\}$, $t \in \mathbb{R}$ onde $\mathrm{erf}(x)$ é a função erro.

DISTRIBUIÇÃO NORMAL (OU GAUSSIANA)

Uma variável aleatória X tem distribuição normal de parâmetros (μ, σ^2) com $\mu \in \mathbb{R}$ e $\sigma > 0$ se tiver função densidade de probabilidade

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

denotando-se $\mathrm{X} \sim N(\mu, \sigma^2)$. Quando $\mu = 0$ e $\sigma^2 = 1$ dizemos que X tem **distribuição normal standard**.

Propriedades:

1. $\mathbb{E}(\mathrm{X}) = \mu$.

2. $\mathbb{V}(\mathrm{X}) = \sigma^2$.

3. $\mathbb{E}\big[\mathrm{X} - \mathbb{E}(\mathrm{X})\big]^k = \begin{cases} 1 \cdot 3 \cdot \ldots \cdot (k-1)\sigma^k & \text{se} \quad k \text{ é par} \\ \\ 0 & \text{se} \quad k \text{ é ímpar} \end{cases}, k \in \mathbb{N}$.

4. $\mathrm{X} \sim N(\mu, 2\sigma^2) = \mathcal{S}_2(\sigma, *, \mu)$.

5. O valor $x_0 = \mu$ é moda.

6. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = e^{i\mu t - \frac{t^2\sigma^2}{2}}$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO NORMAL MULTIVARIADA

Um vector aleatório $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ tem distribuição normal multivariada se tiver função característica

$$\varphi_{\mathbf{X}}(\boldsymbol{t}) = \exp\left\{ i\, \boldsymbol{t}^T \boldsymbol{\mu} - \frac{1}{2}\, \boldsymbol{t}^T \boldsymbol{C} \boldsymbol{t} \right\}, \quad \boldsymbol{t} \in \mathbb{R}^n$$

onde $\boldsymbol{C} \in \mathcal{M}_{n\times n}(\mathbb{R})$ é uma matriz simétrica definida não-negativa e $\boldsymbol{\mu} \in \mathbb{R}^n$.

Propriedades:

1. $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ e $\boldsymbol{\mu}$ designa-se por **vector médio** de $\mathbf{X}$.
2. Quando a característica da matriz $\boldsymbol{C}$ é n diz-se que $\mathbf{X}$ tem distribuição normal multivariada não-degenerada designando-se $\boldsymbol{C}$ por **matriz de covariância** e a função de densidade conjunta de $\mathbf{X}$ é,
$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{C})}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \right\}, \quad \boldsymbol{x} \in \mathbb{R}^n.$$

DISTRIBUIÇÃO DE PARETO

Uma variável aleatória X tem distribuição de Pareto com parâmetros (x_0, α) onde $x_0, \alpha > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{\alpha}{x_0}\left(\dfrac{x_0}{x}\right)^{\alpha+1} & \text{se} \quad x > x_0 \\ 0 & \text{se} \quad x \leqslant x_0 \end{cases}$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = \frac{\alpha x_0^k}{\alpha - k}$, $\alpha > k$.
2. $\mathbb{V}(\mathrm{X}) = \frac{\alpha\, x_0^2}{(\alpha-1)(\alpha-2)}$, $\alpha > 2$.
3. O valor x_0 é moda.
4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = -x_0^\alpha t^\alpha \cos\left(\frac{\pi\alpha}{2}\right)\Gamma(1-\alpha) + {}_1\mathrm{F}_2\left(-\frac{\alpha}{2}; \frac{1}{2}, 1-\frac{\alpha}{2}; -\frac{\alpha^2 t^2}{4}\right) +$
$+\frac{1}{\alpha-1}\left[x_0 t i \alpha\, {}_1\mathrm{F}_2\left(\frac{1}{2}-\frac{\alpha}{2}; \frac{3}{2}, \frac{3}{2}-\frac{\alpha}{2}; -\frac{x_0^2 t^2}{4}\right)\operatorname{sign}(t)\right] + i x_0^\alpha t^\alpha \Gamma(1-\alpha)\operatorname{sign}(t)\sin\left(\frac{\pi\alpha}{2}\right)$, $t \in \mathbb{R}$ onde ${}_p\mathrm{F}_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x)$ é a função hipergeométrica generalizada.[8]

DISTRIBUIÇÃO FUNÇÃO POTÊNCIA

Uma variável aleatória X tem distribuição função potência com parâmetros (x_0, α) onde $x_0, \alpha > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{\alpha x^{\alpha-1}}{x_0^\alpha} & \text{se} \quad x \in [0, x_0] \\ 0 & \text{se} \quad x \notin [0, x_0] \end{cases}$$

Propriedades:

[8] ${}_p\mathrm{F}_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x) = \displaystyle\sum_{n=0}^{\infty} \frac{(a_1)_n \ldots (a_p)_n}{(b_1)_n \ldots (b_q)_n} \frac{x^n}{n!}$, $x \in \mathbb{R}$ onde $(\alpha)_n := \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}$ é o símbolo de Pochhammer.

1. $\mathbb{E}\left(\mathrm{X}^k\right) = \frac{\alpha x_0^k}{\alpha+k}$.

2. $\mathbb{V}(\mathrm{X}) = \frac{\alpha\, x_0^2}{(\alpha+1)^2(\alpha+2)}$, $\alpha > 2$.

3. O valor x_0 é moda quando $\alpha > 1$.

4.
$$\mathbb{E}\big(e^{it\mathrm{X}}\big) = {}_1\mathrm{F}_2\left(-\frac{\alpha}{2}; \frac{1}{2}, 1+\frac{\alpha}{2}; -\frac{\alpha^2 t^2}{4}\right) + \\ + \frac{1}{\alpha+1}\left(x_0 t i \alpha\, {}_1\mathrm{F}_2\left(\frac{1}{2}+\frac{\alpha}{2}; \frac{3}{2}, \frac{3}{2}+\frac{\alpha}{2}; -\frac{x_0^2 t^2}{4}\right)\mathrm{sign}(t)\right), \quad t \in \mathbb{R}.$$

DISTRIBUIÇÃO χ.

Uma variável aleatória X tem distribuição χ com $\nu > 0$ graus de liberdade se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)} x^{\nu-1} e^{-\frac{x^2}{2}} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

e denota-se $\mathrm{X} \sim \chi(\nu)$.

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = \frac{2^{\frac{k}{2}}\Gamma\left(\frac{\nu+k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = \nu - 2\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right]^2$.

3. O valor $x_0 = \sqrt{\nu - 1}$ é a única moda.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} \sum_{n=0}^{\infty} \frac{\left(i\sqrt{2}t\right)^n}{n!} \Gamma\left(\frac{\nu+n}{2}\right)$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO χ^2

Uma variável aleatória X tem distribuição χ^2 com ν graus de liberdade se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

e denota-se $\mathrm{X} \sim \chi^2(\nu)$.

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = \nu(\nu+2)\ldots\big[\nu + 2(k-1)\big]$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = 2\nu$.

3. O valor $x_0 = \nu - 2$ é moda quando $\nu > 2$; para $\nu = 2$ a moda é 2.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = (1-2it)^{-\frac{\nu}{2}}$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO χ^2 NÃO-CENTRAL

Uma variável aleatória X tem distribuição χ^2 não-central com parâmetro de não-centralidade λ e ν graus de liberdade se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{e^{-\frac{x+\lambda}{2}} x^{\frac{\nu}{2}+1}}{2^{\frac{\nu}{2}}} \displaystyle\sum_{n=0}^{\infty} \dfrac{(\lambda x)^n}{n!\, 2^{2n}\Gamma\left(n+\frac{\nu}{2}\right)} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

e denota-se $\mathrm{X} \sim \chi^2(\nu, \lambda)$.

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = 2^k e^{-\frac{\lambda}{2}} \Gamma\left(k+\frac{\nu}{2}\right) {}_1\widetilde{\mathrm{F}}_1\left(k+\frac{\nu}{2}, \frac{\nu}{2}, \frac{\lambda}{2}\right)$, $k \in \mathbb{N}$ onde

$${}_p\widetilde{\mathrm{F}}_q(a_1,\ldots,a_p;b_1,\ldots,b_q;x) = \frac{{}_p\mathrm{F}_q(a_1,\ldots,a_p;b_1,\ldots,b_q;x)}{\Gamma(b_1)\ldots\Gamma(b_q)}$$

é a função hipergeométrica generalizada regularizada.

2. $\mathbb{V}(\mathrm{X}) = 2(2\lambda + \nu)$.

DISTRIBUIÇÃO DE RAYLEIGH

Uma variável aleatória X tem distribuição de Rayleigh com parâmetro $\sigma > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} & \text{se} \quad x > 0 \\ \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

Propriedades:

1. $\mathbb{E}\left(\mathrm{X}^k\right) = 2^{\frac{k}{2}} \sigma^k \Gamma\left(\frac{k}{2}+1\right)$.

2. $\mathbb{V}(\mathrm{X}) = \sigma^2\left(2 - \frac{\pi}{2}\right)$.

3. O valor $x_0 = \sigma$ é moda.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = 1 + ie^{-\frac{\sigma^2 t^2}{2}} \sqrt{\frac{\pi}{2}}\sigma t \left[1 - \operatorname{erf}\left(\frac{i\sigma t}{\sqrt{2}}\right)\right]$, $t \in \mathbb{R}$.

DISTRIBUIÇÃO DE SNEDECOR

Uma variável aleatória X tem distribuição de Snedecor com (ν_1, ν_2) graus de liberdade se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\nu_1^{\frac{\nu_1}{2}}\nu_2^{\frac{\nu_2}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} x^{\frac{\nu_1}{2}-1}(\nu_2+\nu_1 x)^{-\frac{\nu_1+\nu_2}{2}} & \text{se} \quad x > 0 \\ 0 & \text{se} \quad x \leqslant 0 \end{cases} \qquad (\nu_1, \nu_2 > 0)$$

Propriedades:

1. $\mathbb{E}(\mathrm{X}^k) = \nu_2^k \frac{\Gamma\left(\frac{\nu_1}{2}+k\right)\Gamma\left(\frac{\nu_2}{2}-k\right)}{\nu_1^k \Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}$, $\nu_2 > 2k$.
2. $\mathbb{V}(\mathrm{X}) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ $(\nu_2 > 4)$.
3. O valor $x_0 = \frac{(\nu_1-2)\nu_2}{\nu_1(\nu_2+2)}$ é a única moda quando $\nu_1 > 2$.
4. Se X e Y forem duas variáveis aleatórias independentes admitindo distribuição χ^2 com, respectivamente, n e m graus de liberdade então a variável aleatória $\dfrac{\mathrm{X}/n}{\mathrm{Y}/m}$ tem distribuição de Snedecor com (n, m) graus de liberdade.

DISTRIBUIÇÃO DE SNEDECOR NÃO-CENTRAL

Uma variável aleatória X tem distribuição de Snedecor de parâmetros (ν_1, ν_2, λ) com $\nu_1, \nu_2 \in \mathbb{N}$ e $\lambda > 0$ se tiver função densidade de probabilidade

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\nu_1^{\frac{\nu_1}{2}}\nu_2^{\frac{\nu_2}{2}}e^{-\frac{\lambda}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} x^{\frac{\nu_1}{2}-1}(\nu_2+\nu_1 x)^{-\frac{\nu_1+\nu_2}{2}} {}_1\mathrm{F}_1\left[\dfrac{\nu_1+\nu_2}{2}, \dfrac{\nu_1}{2}, \dfrac{\nu_1\lambda x}{2(\nu_1 x+\nu_2)}\right] & \text{se} \quad x > 0 \\ 0 & \text{se} \quad x \leqslant 0 \end{cases}$$

onde ${}_p\mathrm{F}_q$ é a função hipergeométrica generalizada.

Propriedades:

1. $\mathbb{E}(\mathrm{X}) = \frac{\nu_2(\nu_1+\lambda)}{\nu_1(\nu_2-2)}$, $\nu_2 > 2$.
2. $\mathbb{V}(\mathrm{X}) = \frac{2\nu_2^2\left((\nu_1+\lambda)^2+(\nu_2-2)(\nu_1+2\lambda)\right)}{\nu_1^2(\nu_2-2)^2(\nu_2-4)}$, $\nu_2 > 4$.

DISTRIBUIÇÃO t

Uma variável aleatória X tem distribuição t de Student com $\nu > 0$ graus de liberdade se tiver função densidade de probabilidade

$$f(x) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}.$$

Propriedades:

1. $$\mathbb{E}\big(\mathrm{X}^k\big) = \begin{cases} \dfrac{\nu^{\frac{k}{2}}}{\sqrt{\pi}} \dfrac{\Gamma\big(\frac{\nu+k}{2}\big)\Gamma\big(\frac{k+1}{2}\big)}{\Gamma\big(\frac{\nu}{2}\big)} & \text{se} \quad k \text{ é par} \\ 0 & \text{se} \quad k \text{ é ímpar} \end{cases}, \; k \in \mathbb{N},\; \frac{\nu}{2} > 2.$$

2. $\mathbb{V}(\mathrm{X}) = \frac{\nu}{\nu-2}$, $\nu > 2$.

3. O valor $x_0 = 0$ é moda.

4. $\mathbb{E}\big(e^{it\mathrm{X}}\big) = \dfrac{2^{1-\frac{\nu}{2}}\nu^{\frac{\nu}{4}}|t|^{\frac{\nu}{2}}\mathrm{K}_{\frac{\nu}{2}}\big(\sqrt{\nu}|t|\big)}{\Gamma\big(\frac{\nu}{2}\big)}$, $t \in \mathbb{R}$ onde $\mathrm{K}_n(x)$ é a função de Bessel modificada.[9]

DISTRIBUIÇÃO *t* NÃO-CENTRAL

Uma variável aleatória X tem distribuição t de Student não-central com parâmetro de não-centralidade $\lambda \in \mathbb{R}$ e $\nu > 0$ graus de liberdade se tiver função densidade de probabilidade

$$f(x) = \frac{\nu^{\frac{\nu}{2}} e^{-\frac{\lambda^2}{2}}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)\left(\nu + x^2\right)^{\frac{\nu+1}{2}}} \sum_{j=0}^{\infty} \Gamma\left(\frac{\nu+j+1}{2}\right)\left(\frac{\lambda^j}{j!}\right)\left(\frac{2x^2}{\nu+x^2}\right)^{\frac{j}{2}}, \quad x \in \mathbb{R}.$$

Propriedades:

1. $$\mathbb{E}\big(\mathrm{X}^k\big) = \begin{cases} \dfrac{\nu^{\frac{\nu}{2}}\Gamma\big(\frac{\nu-k}{2}\big)}{2^{\frac{\nu}{2}}\Gamma\big(\frac{\nu}{2}\big)} \sum_{j=0}^{k} \dfrac{(2k)!\lambda^{2j}}{(2j)!(k-j)!2^{k-j}} & \text{se} \quad k \text{ é par} \\ \dfrac{\nu^{\frac{\nu}{2}}\Gamma\big(\frac{\nu-k}{2}\big)}{2^{\frac{\nu}{2}}\Gamma\big(\frac{\nu}{2}\big)} \sum_{j=1}^{k} \dfrac{(2k-1)!\lambda^{2j-1}}{(2j-1)!(k-j)!2^{k-j}} & \text{se} \quad k \text{ é ímpar} \end{cases}, \; k \in \mathbb{N},\; \nu > k.$$

DISTRIBUIÇÃO *t* MULTIVARIADA

Um vector aleatório $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ tem distribuição t multivariada com k graus de liberdade, vector de desvio $\boldsymbol{\mu}$ e matriz de precisão $\boldsymbol{T}$ se tiver função de densidade conjunta

$$f(\boldsymbol{x}) = \frac{\Gamma\left(\frac{n+k}{2}\right)\sqrt{\det \boldsymbol{T}}}{\Gamma\left(\frac{k}{2}\right)(k\pi)^{\frac{n}{2}}}\left[1 + \frac{1}{k}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{T}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{-\frac{n+k}{2}}, \quad \boldsymbol{x} \in \mathbb{R}^n$$

onde $\boldsymbol{T} \in \mathcal{M}_{n\times n}(\mathbb{R})$ é uma matriz simétrica definida positiva.

Propriedades:

1. $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$.

2. $\mathbb{E}\big[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\big] = \frac{k}{k-1}\boldsymbol{T}^{-1}$, $k > 2$.

---

[9] $$\mathrm{K}_n(x) = (-1)^{n+1} i^{-n} \mathcal{B}_n(ix)\log\left(\frac{x}{2}\right) + \frac{1}{2}\left(\frac{2}{x}\right)^n \sum_{m=0}^{n-1} \frac{(n-m-1)!}{m!}\left(-\frac{x^2}{4}\right)^m +$$

$$+\frac{(-1)^n}{2}\left(\frac{x}{2}\right)^n \sum_{m=0}^{\infty}\left[\frac{\Gamma'(m+1)}{\Gamma(m+1)} + \frac{\Gamma'(n+m+1)}{\Gamma(n+m+1)}\right]\frac{(x/2)^{2m}}{m!(n+m)!}$$

onde $\mathcal{B}_n(z)$ é a função de Bessel.

DISTRIBUIÇÃO UNIFORME

Uma variável aleatória X tem distribuição no intervalo $[a,b]$ $(a<b)$ se tiver função densidade de probabilidade

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{se} \quad x\in[a,b]\\[2ex] 0 & \text{se} \quad x\notin[a,b]\end{cases}$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big)=\frac{b^{k+1}-a^{k+1}}{(b-a)(k-1)}$, $k\in\mathbb{N}$.
2. $\mathbb{V}(\mathrm{X})=\frac{(b-a)^2}{12}$.
3. $\mathbb{E}\big(e^{it\mathrm{X}}\big)=\frac{e^{itb}-e^{ita}}{i(b-a)t}$, $t\in\mathbb{R}$.

DISTRIBUIÇÃO UNIFORME MULTIVARIADA

Um vector aleatório $\mathbf{X}=(\mathrm{X}_1,\ldots,\mathrm{X}_n)$ tem distribuição uniforme multivariada sobre um boreliano B limitado de $\mathbb{R}^n$ tal que $\mathrm{Leb}(B)>0$ se tiver função de densidade conjunta

$$f(x_1,\ldots,x_n)=\begin{cases}\dfrac{1}{\mathrm{Leb}(B)} & \text{se} \quad (x_1,\ldots,x_n)\in B\\[2ex] 0 & \text{se} \quad (x_1,\ldots,x_n)\notin B\end{cases}$$

DISTRIBUIÇÃO DE WEIBULL

Uma variável aleatória X tem distribuição de Weibull com parâmetro $\alpha>0$ se tiver função densidade de probabilidade

$$f(x)=\begin{cases}\alpha(-x)^{\alpha-1}e^{-(-x)^\alpha} & \text{se} \quad x<0\\[2ex] 0 & \text{se} \quad x\geqslant 0\end{cases}$$

Propriedades:

1. $\mathbb{E}\big[\,|\mathrm{X}|^k\,\big]=\Gamma\left(1+\frac{k}{\alpha}\right)$, $k\in\mathbb{N}$.
2. $\mathbb{V}(\mathrm{X})=\frac{2}{\alpha}\Gamma\left(\frac{2}{\alpha}\right)-\frac{1}{\alpha^2}\left[\Gamma\left(\frac{1}{\alpha}\right)\right]^2$.
3. O valor $x_0=-\left(1-\frac{1}{\alpha}\right)^{\frac{1}{\alpha}}$ é moda se $\alpha>1$; o valor $x_0=0$ é moda se $\alpha\leqslant 1$.

DISTRIBUIÇÃO WEIBULL-GNEDENKO

Uma variável aleatória X tem distribuição Weibull-Gnedenko de parâmetros (α,λ) com $\alpha\in\mathbb{R}$ e $\lambda>0$ se tiver função densidade de probabilidade

$$f(x)=\begin{cases}|\alpha|\,\lambda x^{\alpha-1}e^{-\lambda x^\alpha} & \text{se} \quad x>0\\[2ex] 0 & \text{se} \quad x\leqslant 0\end{cases}$$

Propriedades:

1. $\mathbb{E}\big(\mathrm{X}^k\big) = \lambda^{-\frac{k}{\alpha}}\Gamma\left(\frac{k}{\alpha}+1\right)$, $k \in \mathbb{N}$.

2. $\mathbb{V}(\mathrm{X}) = \lambda^{-\frac{2}{\alpha}}\left\{\frac{2}{\alpha}\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha^2}\left[\Gamma\left(\frac{1}{\alpha}\right)\right]^2\right\}$.

3. O valor $x_0 = \left(\frac{\alpha-1}{\alpha\lambda}\right)^{\frac{1}{\alpha}}$ é moda se $\alpha > 1$; o valor $x_0 = 0$ é moda se $\alpha \leqslant 1$.

Bibliografia

[1] Apostol, T. M. (1967). *Calculus - Volume 1* (second edition). John Wiley & Sons.

[2] Apostol, T. M. (1969). *Calculus - Volume 2* (second edition). John Wiley & Sons.

[3] Apostol, T. M. (1976). *Introduction to Analytic Number Theory*. Springer-Verlag.

[4] Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

[5] Anderson, B. D. O. & Moore, J. B. (1975). On martingales and least squares linear system identification. *Technical report* EE7522.

[6] Anderson, B. D. O. & Moore, J. B. (1976). A matrix Kronecker lemma. *Linear Algebra and its Applications* 15, p. 227-234.

[7] Arce, G. R. (2005). *Nonlinear Signal Processes: A Statistical Approach*. John Wiley & Sons.

[8] Arfken, G. (1985). *Mathematical Methods for Physicists* (third edition). Academic Press.

[9] Bates, D. M. & Watts, D. G. (1988). *Nonlinear Regression Analysis and its Applications*. John Wiley & Sons.

[10] Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons.

[11] Bingham, N. H., Goldie, C. M. & Teugels, J. L. (1987). *Regular Variation*. Cambridge University.

[12] Billingsley, P. (1995). *Probability and Measure* (third edition). John Wiley & Sons.

[13] Cambanis, S., Huang, S. & Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11, p. 368-385.

[14] Chang, Der-Shin & Chang, Mei-Rong (1996). A note on hypothesis testing in stochastic regression models. *Stat. Sin.*, Vol. 6, p. 491-498.

[15] Chen, X. (1994). Some results on consistency of LS estimates. *Chin. Sci. Bull.*, Vol. 39(22), p. 1872-1876.

[16] Chen, X. (1995). Consistency of LS estimates of multiple regression under a lower order moment condition. *Sci. Chin.*, Vol. 38(12), p. 1420-1431.

[17] Chen, X. (2001). A note on the consistency of LS estimates in linear models. *Chin. Ann. Math. Ser. B*, Vol. 22(4), p. 471-474.

[18] Chow, Y. S. & Teicher, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales.* Springer.

[19] Chung, K. L. (2001). *A Course in Probability Theory* (third edition). Academic Press.

[20] Conway, J. H. & Guy, R. K. (1996). *The Book of Numbers.* Springer-Verlag.

[21] Dacunha-Castelle, D. et Duflo, M. (1982). *Probabilités et Statistiques: Problèmes à Temps Fixe.* Masson.

[22] Drygas, H. (1971). Consistency of the least squares and Gauss-Markov estimators in regression models. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 17, p. 309-326.

[23] Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression model. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 34, p. 119-127.

[24] Dudley, R. M. (2004). *Real Analysis and Probability.* Cambridge University Press.

[25] Durrett, R. (1996). *Probability: Theory and Examples* (second edition). Wadsworth Publishing Company.

[26] Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events.* Springer.

[27] Fang, K., Kotz, S. & Ng, K. (1990). *Symmetric Multivariate and Related Distributions.* Monographs on Statistics and Applied Probability 36, Chapman & Hall.

[28] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications - Volume I* (third edition). John Wiley & Sons.

[29] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications - Volume II* (second edition). John Wiley & Sons.

[30] Fernandez, P. J. (1976). *Medida e Integração.* Projecto Euclides.

[31] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics.* Robert E. Krieger Publishing Company.

[32] Gnedenko, B. V. & Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading, MA.

[33] Griliches, Z. & Intriligator, M. D. (2001). *Handbook of Econometrics - Volume 1* (sixth edition). North-Holland.

[34] Gujarati, D. N. (2004). *Basic Econometrics*. McGraw-Hill.

[35] Hall, Alastair (1986). Conditions for a matrix Kronecker Lemma. *Linear Algebra and its Applications*, 76:271-277.

[36] Halmos, P. R. (1965). *Measure Theory* (tenth edition). D. Van Nostrand Company, Inc.

[37] Hardy, G., Littlewood, J. E. & Pólia, G. (1952). *Inequalities*. Cambridge University Press.

[38] Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer.

[39] Hewitt, E. & Stromberg, K. (1965). *Real and Abstract Analysis*. Springer.

[40] Horn, R. A. & Johnson, C. R. (1985). *Matrix Analysis*. Cambridge University Press.

[41] Ibragimov, I. A. & Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.

[42] James, B. R. (2004). *Probabilidade: Um Curso de Nível Intermediário* (terceira edição). Projecto Euclides.

[43] Jin, M. (1995). Some new results of the strong consistency of multiple regression coefficients. *Proceedings of the Second Asian Mathematical Conference 1995* (Tangmanee, S. & Schulz, E. eds.), p. 514-519, World Scientific.

[44] Jin, M. & Chen, X. (1999). Strong consistency of least squares estimate in multiple regression when the error variance is infinite. *Stat. Sin.*, Vol. 9(1), p. 289-296.

[45] Johnston, J. (1991). *Econometric Methods* (third edition). McGraw-Hill International Editions.

[46] Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. & Lee, T. C. (1988). *Introduction to the Theory and Practice of Econometrics* (second edition). John Wiley & Sons.

[47] Kolmogorov, A. N. (1956). *Foundations of the Theory of Probability* (second edition). Chelsea Publishing Company, New York.

[48] Kolmogorov, A. N. e Fomin, S. V. (1982). *Elementos da Teoria das Funções e de Análise Funcional*. Mir.

[49] Koroliouk, V., Portenko, N., Skorokhod, A. et Tourbine, A. (1983). *Aide-Mémoire de Théorie des Probabilités et de Statistique Mathématique*. Mir.

[50] Kudriávtsev, L. D. (1983). *Curso de Análisis Matemático* (*Tomo 1*). Mir.

[51] Kudriávtsev, L. D. (1984). *Curso de Análisis Matemático* (*Tomo 2*). Mir.

[52] Lai, T. L. & Robbins, H (1977). Strong consistency of least-squares estimates in regression models. *Proc. Natl. Acad. Sci. USA* Vol. 74(7), p. 2667-2669.

[53] Lai, T. L., Robbins, H & Wei, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Natl. Acad. Sci. USA* Vol. 75(7), p. 3034-3036.

[54] Lai, T. L., Robbins, H & Wei, C. Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multivariate Anal.* 9, p. 343-362.

[55] Loève, M. (1977). *Probability Theory I* (fourth edition). Springer.

[56] Loève, M. (1978). *Probability Theory II* (fourth edition). Springer.

[57] Lukacs, E. (1975). *Stochastic Convergence* (second edition). Academic Press.

[58] Machado, A. (1976). *Medida e Integração*. Textos e Notas do CMAF, 3.

[59] Magalhães, L. T. (1992). *Álgebra Linear como Introdução à Matemática Aplicada*. Texto Editora.

[60] Makarov, B. M., Goluzina, M. G., Lodkin, A. A. & Podkorytov, A. N. (1992). *Selected Problems in Real Analysis*. American Mathematical Society.

[61] Mexia, J. T. (1987). Multitreatment regression designs. *Trabalhos de Investigação da FCT-UNL*, 1.

[62] Mexia, J. T. & Corte Real, P. (2001). Extension of Kolmogorov's strong law to multiple regression. *Revista de Estatística*, N. 24 (2º quadrimestre de 2001), p. 277-278.

[63] Mexia, J. T., Corte Real, P., Esquível, M. L. e Lita da Silva, J. (2005). Convergência do estimador dos mínimos quadrados em modelos lineares. *Estatística Jubilar. Actas do XII Congresso da Sociedade Portuguesa de Estatística*, Edições SPE, p. 455-466.

[64] Mexia, J. T., Corte Real, P., Esquível, M. L. e Lita da Silva, J. (2006). Convergência forte de estimadores, erros com decaimento exponencial no infinito e colapso de erros radiais. *Ciência Estatística. Actas do XIII Congresso Anual da Sociedade Portuguesa de Estatística*, Edições SPE, p. 447-458.

[65] Mexia, J. T. & Lita da Silva, J. (2006). Least squares estimator consistency: a geometric approach. *Discussiones Mathematicae - Probability and Statistics*, 26, p. 19-45.

[66] Mexia, J. T. e Lita da Silva, J. (2006). A consistência do estimador dos mínimos quadrados em domínios de atracção maximais. *Ciência Estatística. Actas do XIII Congresso Anual da Sociedade Portuguesa de Estatística*, Edições SPE, p. 481-492.

[67] Mexia, J. T. & Lita da Silva, J. (2006). Sufficient conditions for the strong consistency of least squares estimator with α-stable errors. (submetido para publicação).

[68] Monteiro, A. J. A. (1989). *Álgebra Linear e Geometria Analítica*. Edição da Associação dos Estudantes da FCUL.

[69] Moreira, E. E., Ribeiro, A. B., Mateas, E. P., Mexia, J. T. & Ottosen, L. M. (2005). Regression modeling of electrodialytic removal of Cu, Cr, and As fro CCA treated timber waste. *Wood Science and Technology* (to appear).

[70] Murteira, B. (1990). *Probabilidades e Estatística - Volume 1* (segunda edição). McGraw-Hill.

[71] Murteira, B. (1990). *Probabilidades e Estatística - Volume 2* (segunda edição). McGraw-Hill.

[72] Murteira, Bento, Ribeiro, C. S., Andrade e Silva, J. e Pimenta, C. (2002). *Introdução à Estatística*. McGraw-Hill.

[73] Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes* (third edition). McGraw-Hill International Editions.

[74] Pestman, W. (1998). *Mathematical Statistics*. Walter de Gruyter Berlin.

[75] Port, S. C. (1994). *Theoretical Probability for Applications*. John Wiley & Sons.

[76] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (second edition). John Wiley & Sons.

[77] Reiss, R. D. & Thomas M. (2001). *Statistical Analysis of Extreme Values*. Birkhäuser.

[78] Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer.

[79] Resnick, S. I. (1999). *A Probability Path*. Birkhäuser Boston.

[80] Ribeiro, A. B. & Mexia, J. T. (1997). A dynamical model for the electrokinetic removal of copper from a polluted soil. *Journal of Hazardous Materials*, Vol. 56, p. 257-271.

[81] Rohatgi, V. K. & Saleh, A. K. Md. E. (2001). *An Introduction to Probability Theory and Mathematical Statistics* (second edition). John Wiley & Sons.

[82] Samorodnitsky, G. & Taqqu, M. S. (1994). *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall.

[83] Schott, J. R. (1997). *Matrix Analysis for Statistics*. John Wiley & Sons.

[84] Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research*, 55:301-327.

[85] Shiryaev, A. N. (1996). *Probability* (second edition). Springer.

[86] Stoyanov, J. M. (1997). *Counterexamples in Probability* (second edition). John Wiley & Sons.

[87] Struik, D. J. (1987). *História Concisa das Matemáticas*. Ciência Aberta, Gradiva.

[88] Tong, Y. L. (1990). *The Multivariate Normal Distribution*. Springer-Verlag New York Inc.

[89] Uchaikin V. V. & Zolotarev, V. M. (1999). *Chance and Stability. Stable Distributions and Their Applications*. Ultrech.

[90] Wei, C. Z. (1985). Asymptotic properties of least-squares estimates in stochastic regression models. *The Annals of Statistics*, Vol.13(4), p. 1498-1508.

[91] Williams, D. (2001). *Probability with Martingales* (fifth edition). Cambridge University Press.

[92] Yasuda, K. (1996). Additive processes on local fields. *J. Math. Sci. Univ. Tokyo* 3, No. 3, p. 629-654.

[93] Zolotarev, V. M. (1986). *One-Dimensional Stable Distributions*. American Mathematical Society, Providence, R.I.

[94] Zwillinger, D. & Kokoska, S. (2000). *CRC Standard Probability and Statistics Tables and Formulae*. Chapman & Hall/CRC.

Índice de Autores

Índice Remissivo